

Implementasi Algoritma K-Nearest Neighbor dalam Kategori Penyakit Anemia

Wulan Mauliawati¹, Ina Najiyyah²

^{1,2}Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya

e-mail: ¹wulanmauliawati1@gmail.com, ²inajiyah@ars.ac.id

Abstrak

Penelitian ini mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan jenis penyakit anemia berdasarkan indikator fisik dan parameter medis. Data terdiri dari 1.281 entri dari Kaggle dengan fitur-fitur relevan untuk klasifikasi penyakit anemia. Proses meliputi pengumpulan data, pengolahan, seleksi fitur, pengujian model, dan laporan kinerja. Seleksi fitur dilakukan untuk meningkatkan akurasi dan mengurangi kompleksitas model. Hasil eksperimen menunjukkan bahwa KNN mampu mengklasifikasikan jenis anemia dengan akurasi tinggi, terbukti dengan nilai presisi, *recall*, dan F-Measure yang signifikan pada nilai K tertentu. Penelitian ini bertujuan untuk mengembangkan metode klasifikasi penyakit anemia yang lebih efektif dan dapat menjadi referensi untuk penelitian lanjutan di bidang ini.

Kata kunci: K-Nearest Neighbor, Kategori Anemia, Seleksi Fitur, Akurasi Model

Abstract

This research implements the K-Nearest Neighbor (KNN) algorithm to classify types of anemia based on physical indicators and medical parameters. The dataset consists of 1,281 entries from Kaggle with relevant features for anemia classification. The process includes data collection, data processing, feature selection, model testing, and performance reporting. Feature selection was conducted to enhance accuracy and reduce model complexity. The experimental results show that KNN can classify types of anemia with high accuracy, as evidenced by significant precision, recall, and F-measure values at specific K values. This research aims to develop a more effective method for anemia classification and can serve as a reference for further research in this field.

Keywords: K-Nearest Neighbor, Anemia Category, Feature Selection, Model Accuracy

Corresponding Author:

Ina Najiyyah,

Email: inajiyah@ars.ac.id

1. PENDAHULUAN

Organisasi Kesehatan Dunia (WHO) pada tahun 2019, prevalensi global anemia pada wanita usia subur mencapai 29,9% menurut WHO. Di Indonesia, prevalensi anemia pada remaja putri (15-24 tahun) adalah 27,2%, sedangkan pada remaja putra 20,3%. Anemia menjadi masalah kesehatan utama, terutama bagi remaja putri, dengan penyebab utama adalah kebiasaan makan yang kurang optimal dan kurangnya aktivitas fisik [1]. Anemia adalah masalah kesehatan utama di seluruh dunia, terutama di negara berkembang. Kondisi ini ditandai dengan penurunan jumlah sel darah merah, kadar hemoglobin, dan hematokrit, yang mengakibatkan tubuh menerima oksigen lebih sedikit. Jenis anemia dapat diklasifikasikan berdasarkan penyebabnya: penurunan pembentukan sel darah merah, kehilangan sel darah merah, dan peningkatan pemecahan sel darah merah (hemolisis) [2]. Perkembangan teknologi, terutama machine learning, telah membawa manfaat signifikan di bidang pendidikan dan kesehatan. Machine learning memungkinkan mesin belajar mandiri tanpa instruksi langsung. Dengan membandingkan berbagai algoritma, kita dapat menemukan algoritma yang paling akurat dan sesuai untuk konteks tertentu

[3].

Algoritma K-*Nearest Neighbors* (K-NN) adalah teknik klasifikasi yang menentukan kelas data uji berdasarkan kedekatannya dengan data latih terdekat. Dengan mengukur jarak antar objek dan mempertimbangkan k tetangga terdekat, K-NN mengklasifikasikan data baru menggunakan label dari tetangga terdekatnya. Teknik ini diterapkan di berbagai bidang seperti kedokteran, keuangan, dan *text mining* [4]. Penelitian "Penerapan Algoritma Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah" menunjukkan bahwa metode FKNN memberikan hasil optimal dengan akurasi tertinggi 85,01% pada K=13, menggunakan 70% data untuk pelatihan dan 30% untuk pengujian [5]. Siti Julaisha dan rekan-rekannya [6] menggunakan K-NN untuk mengklasifikasikan calon penerima Bidikmisi, menghasilkan akurasi 83,13% pada K=5. Cholil [7] menggunakan K-NN untuk klasifikasi seleksi beasiswa, dengan akurasi rata-rata 90,5% berdasarkan Confusion Matrix. Aldi Tangkelayuk [8] mengevaluasi klasifikasi kualitas air menggunakan K-NN, NBC, dan Decision Tree, di mana K-NN mencapai akurasi tertinggi 86,88%.

Studi ini menggunakan *dataset* untuk klasifikasi penyakit anemia yang diambil dari Kaggle, dengan skor kegunaan (usability) 9,41, menunjukkan kualitas tinggi dan kemudahan penggunaan untuk penelitian. *Dataset* terdiri dari 1.281 entri yang mencakup fitur relevan untuk klasifikasi anemia. Data ini akan diolah dan dianalisis menggunakan algoritma K-*Nearest Neighbor* untuk menentukan penyakit anemia berdasarkan indikator fisik dan parameter medis. Penelitian ini bertujuan untuk menghasilkan metode klasifikasi yang lebih akurat dan efektif. Judul penelitian ini adalah "Implementasi Algoritma K-*Nearest Neighbor* dalam Klasifikasi Penyakit Anemia."

2. TINJAUAN PUSTAKA

2.1. Machine Learning

Machine learning adalah cabang kecerdasan buatan yang mengembangkan algoritma untuk memungkinkan komputer belajar dari data, membuat prediksi, dan mengambil keputusan berdasarkan pola dalam data. Penerapan *machine learning* dalam pelayanan karsipan memiliki potensi besar untuk meningkatkan mutu dan efisiensi pemrosesan dokumen serta klasifikasi arsip administrasi [9]. *Machine learning* dibagi menjadi tiga kategori utama:

1. *Supervised Learning*: Proses pembelajaran diawasi, dengan fokus pada klasifikasi dan regresi.
2. *Unsupervised Learning*: Proses pembelajaran tidak diawasi, termasuk pengelompokan dan reduksi dimensionalitas.
3. *Reinforcement Learning*: Agen perangkat lunak menentukan perilaku ideal untuk memaksimalkan kinerja, diterapkan dalam pengambilan keputusan real-time dan kecerdasan buatan permainan [10].

2.2. Data Mining

Data mining adalah proses analisis dan eksplorasi data dalam skala besar untuk mengidentifikasi pola dan tren tersembunyi, serta mendapatkan informasi yang akurat dan berguna. Ini merupakan bagian dari kecerdasan buatan yang berfokus pada ekstraksi pola dan transformasi data menjadi informasi bermanfaat. *Data mining* dikenal sejak tahun 1990-an akibat peningkatan volume data dalam basis data [11]. Proses data mining menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi informasi dari database besar dan merupakan bagian dari KDD (*Knowledge Discovery in Databases*), yang meliputi tahapan seperti pemilihan data, pra-pengolahan, transformasi, *data mining*, dan evaluasi hasil [12].

2.3. Anemia

Anemia adalah kondisi di mana massa eritrosit (sel darah merah) menurun,

mengakibatkan ketidakmampuan sel darah merah memenuhi kebutuhan oksigen jaringan. Identifikasi anemia umumnya dilakukan melalui penurunan kadar hemoglobin, hematokrit, atau jumlah eritrosit, dengan kadar hemoglobin sebagai indikator utama. Gejala klinis meliputi kulit pucat, penurunan kinerja fisik, dan daya tahan tubuh berkurang. Penyebab anemia termasuk penyerapan zat besi yang buruk, pendarahan, kerusakan sel darah, atau produksi sel darah merah yang terhambat. Pengobatan bisa meliputi obat, suplemen zat besi, perubahan pola makan, transfusi darah, dan penanganan penyakit penyebab anemia. Anemia pada remaja putri dapat memiliki berbagai penyebab yang mempengaruhi kesehatannya [13].

2.4. Klasifikasi

Menurut *World Health Organization* (WHO), anemia adalah kondisi di mana jumlah sel darah merah atau hemoglobin dalam tubuh menurun, menyebabkan volume darah merah berkurang dan jumlah sel darah merah berada di bawah normal. Hemoglobin dalam sel darah merah berfungsi membawa oksigen dari jantung dan paru-paru ke seluruh tubuh. Anemia dapat diklasifikasikan menjadi tiga jenis utama [14]:

1. Anemia Hemolitik: Terjadi karena percepatan penghancuran sel darah merah di pusat perifer.
 2. Anemia Hipoproliferatif (*Aregeneratif*): Disebabkan oleh penurunan produksi sel darah merah sentral dan/atau gangguan pelepasan dari sumsum tulang.
 3. Anemia Hemoragik: Disebabkan oleh kebocoran dari integritas pembuluh darah tubuh, mengakibatkan penurunan massa eritrosit secara akut atau kronis

2.5. Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (K-NN) mengklasifikasikan objek baru berdasarkan tetangga terdekatnya, dengan kelas yang paling umum sebagai hasil klasifikasinya. K-NN unggul dalam menyelesaikan masalah *multiclass*, meskipun sering mengalami kesulitan menemukan tetangga terdekat pada titik *query* dari *dataset* yang digunakan. K-NN melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) antara data, yang biasanya diukur menggunakan jarak *Euclidean* berikut:

1. Menentukan parameter K
 2. Menghitung jarak antara data *training* dan data *testing*

Perhitungan jarak yang paling umum dipakai pada perhitungan pada algoritma KNN adalah menggunakan perhitungan jarak *Euclidean*. Rumusannya adalah sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots \quad (2.1)$$

Dimana:

$d(p,q) = \text{Jarak}$

$\pi_i = \text{sample data} / \text{data training}$

$qi = \text{data uji} / \text{data testing}$

i = variabel data n = dimensi data

3. Mengurutkan jarak yang terbentuk
 4. Menentukan jarak terdekat sampai urutan K
 5. Memasangkan kelas yang bersesuaian
 6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

2.6. RapidMiner

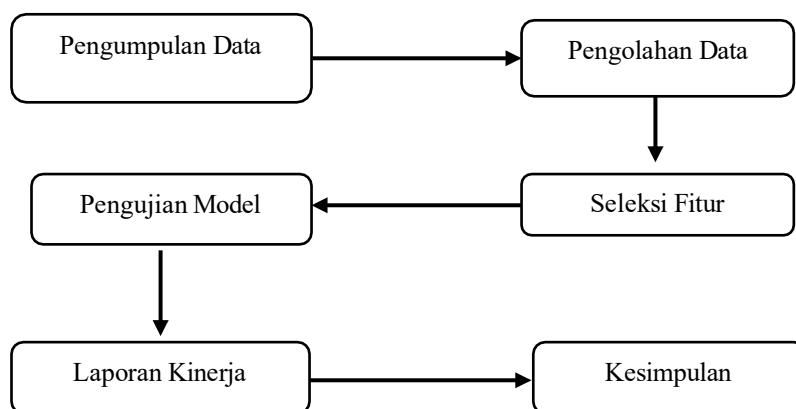
RapidMiner adalah perangkat lunak *open-source* yang menyediakan berbagai alat untuk analisis prediktif dan deskriptif. Lebih dari 500 profesional *machine learning* berkontribusi dalam pengembangannya, mencakup aspek *input* data, *output*, *preprocessing* data, dan visualisasi. *RapidMiner* dapat digunakan sebagai perangkat lunak mandiri atau diintegrasikan ke dalam produk lain untuk analisis data. Dikembangkan dengan Java, *RapidMiner* beroperasi dengan baik di berbagai sistem operasi [15].

3. METODELOGI PENELITIAN

3.1. Tahapan Penelitian

Penelitian atau alat yang digunakan untuk mengumpulkan data dengan teknik penelitian khusus yang meliputi teknik pengumpulan data (observasi, wawancara, dan perekaman audio) serta teknik analisis data (korelasi kuantitatif, statistik) [16].

Penelitian ini mengikuti sesuai tahapan seperti pada gambar dibawah ini.



Gambar 1. Tahapan Penelitian

3.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle, berjudul "Anemia Classification Data," yang relevan dengan topik klasifikasi jenis anemia. Dataset ini dipilih karena memenuhi kriteria kelengkapan data dan minimnya nilai yang hilang. Dataset mencakup berbagai fitur medis dari pasien seperti jumlah sel darah putih, persentase limfosit dan neutrofil, jumlah limfosit dan neutrofil, jumlah sel darah merah, hemoglobin, hematokrit, volume rata-rata eritrosit, rata hemoglobin per eritrosit, konsentrasi hemoglobin rata dalam eritrosit, jumlah trombosit, distribusi ukuran trombosit, dan prokalsitonin. Fitur-fitur ini penting untuk mendiagnosis anemia berdasarkan parameter CBC. Berikut adalah tabel yang merinci proses pengumpulan data:

Tabel 1. Data Mentah Penelitian

JSDP	PL	PN	JL	JN	JSM	HB	HTK	VRSM	RHPSDM	KHRSR	JT	DUT	PK	JA
10	43.2	50.1	4.3	5	2.77	7.3	24.2	87.7	26.3	30.1	189	12.5	0.17	Anemia hipokromik nomosistik
10	42.4	52.3	4.2	5.3	2.84	7.3	25	88.2	25.7	20.2	180	12.5	0.16	Anemia hipokromik nomosistik
7.2	30.7	60.7	2.2	4.4	3.97	9	30.5	77	22.6	29.5	148	14.3	0.14	Anemia defisiensi besi

Tabel diatas menunjukkan data mentah penelitian yang diambil dari *dataset* yang diunduh dari Kaggle. Penelitian ini menggunakan 1,281 data dengan 15 atribut, ditambah 1 atribut spesial yang mencakup 9 kelas, antara lain Anemia hipokromik normosistik, Anemia defisiensi besi, Anemia mikrositik, Anemia makrositik, Leukemia, sehat, Trombositopenia, Leukemia dengan Trombositopenia, dan Anemia normokromik normosistik. Namun, hanya 10 *dataset* yang dicantumkan sebagai contoh.

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil dalam bab ini disusun berdasarkan hasil penelitian yang dilakukan. Penelitian ini menggunakan platform *RapidMiner* dengan menerapkan algoritma K-Nearest Neighbor (KNN) untuk menguji data yang telah dianalisis. Berbagai komponen *RapidMiner* dan proses pengolahan data yang diimplementasikan juga digunakan.

4.1.1 Perhitungan Algoritma K-Nearest Neighbor

Berikut ini adalah rekapitulasi data penelitian yang digunakan pada tabel IV.1 dengan keterangan sebagai berikut: Jumlah sel darah putih = JS DP, Persentase limfosit = PL, Persentasi neutrofil = PN, Jumlah limfosit = JL, Jumlah neutrofil = JN, Jumlah sel darah merah = JSM, Hemoglobin = HB, Hematokrit = HTK, Volume Rata-rata Sel Darah Merah = VRSM, Rata-rata Hemoglobin per Sel Darah Merah = RH PSDM, Konsentrasi Hemoglobin Rata-rata dalam Sel Darah Merah = KHR SR, Jumlah Trombosit = JT, Distribusi Ukuran Trombosit = DUT, Prokalsitonin = PK, Jenis Anemia berdasarkan parameter CBC = JA

Tabel 2. Data Penelitian

JS DP	PL	PN	JL	JN	JSM	HB	HTK	VRSM	RH PSDM	KHR SR	JT	DUT	PK	JA
10	43.2	50.1	4.3	5	2.77	7.3	24.2	87.7	26.3	30.1	189	12.5	0.17	Anemia hipokromik normositik
10	42.4	52.3	4.2	5.3	2.84	7.3	25	88.2	25.7	20.2	180	12.5	0.16	Anemia hipokromik normositik
7.2	30.7	60.7	2.2	4.4	3.97	9	30.5	77	22.6	29.5	148	14.3	0.14	Anemia defisiensi besi

Berikut adalah tabel yang memisahkan perhitungan jarak *Euclidean* antara baris 1 dan baris 2 dsb dari dataset Perhitungan Manual Jarak *Euclidean*:

1. Baris 1 dengan baris 2:

Tabel 3. Baris 1 Dengan Baris 2

Baris	JS DP	PL	PN	JL	JN	JSM	HB	HTK	VRSM	RH PSDM	KHR SR	JT	DUT	PK
1	10	43.2	50.1	4.3	5	2.77	7.3	24.2	87.7	26.3	30.1	189	12.5	0.17
2	10	42.4	52.3	4.2	5.3	2.84	7.3	25	88.2	25.7	20.2	180	12.5	0.16

Hasil perhitungan dari baris 1 dengan baris 2 adalah sebagai berikut:

$$d_{1,2} = \sqrt{(10 - 10)^2 + (43.2 - 42.4)^2 + (50.1 - 52.3)^2 + (4.3 - 4.2)^2 + (5 - 5.3)^2 + (2.77 - 2.84)^2 + (7.3 - 7.3)^2 + (24.2 - 25)^2 + (87.7 - 88.2)^2 + (26.3 - 25.7)^2}$$

$$d_{1,2} = \sqrt{0 + 0.64 + 4.84 + 0.01 + 0.09 + 0.0049 + 0 + 0.64 + 0.25 + 0.36}$$

$$d_{1,2} = \sqrt{6.8449}$$

$$d_{1,2} = 2.62$$

2. Baris 1 dengan baris 3:

Tabel 4. Baris 1 Dengan Baris 3

Baris	JS DP	PL	PN	JL	JN	JSM	HB	HTK	VRSM	RH PSDM	KHR SR	JT	DUT	PK
1	10	43.2	50.1	4.3	5	2.77	7.3	24.2	87.7	26.3	30.1	189	12.5	0.17
3	7.2	30.7	60.7	2.2	4.4	3.97	9	30.5	77	22.6	29.5	148	14.3	0.14

Hasil perhitungan dari baris 1 dengan baris 3 adalah sebagai berikut:

$$d_{1,3} = \sqrt{(10 - 7.2)^2 + (43.2 - 30.7)^2 + (50.1 - 60.7)^2 + (4.3 - 2.2)^2 + (5 - 4.4)^2 + (2.77 - 3.97)^2 + (7.3 - 9)^2 + (24.2 - 30.5)^2 + (87.7 - 77)^2 + (26.3 - 22.6)^2}$$

$$d_{1,3} = \sqrt{7.84 + 157.29 + 112.36 + 4.41 + 0.36 + 1.44 + 2.89 + 39.69 + 114.49 + 13.69}$$

$$d_{1,3} = \sqrt{454.61}$$

$$d_{1,3} = 21.31$$

Hasil Akhir Jarak *Euclidean*:

$$\begin{aligned} d_{1,2} &= 2.61 \\ d_{1,3} &= 21.31 \end{aligned}$$

4.1.2 Perhitungan Menggunakan Data

Untuk memahami perhitungan ini, kita memerlukan *Confusion Matrix*, yang merupakan tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi.

1. Anemia hipokromik normositik:

True Positive (TP): 234

False Positive (FP): $14 + 4 + 9 + 17 + 2 + 20 + 1 + 2 = 69$

False Negative (FN): $12 + 1 + 13 + 5 + 13 + 0 + 0 + 0 = 45$

True Negative (TN): $1281 - (TP + FP + FN) = 1281 - (234 + 69 + 45) = 933$

$$\text{AKURASI: } \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(234 + 933)}{(234 + 933 + 69 + 45)} = \frac{(1167)}{(1281)} = 0,911 \text{ atau } 91,1\%$$

$$\text{PRESISI: } \frac{TP}{TP + FP} = \frac{234}{234 + 69} = \frac{234}{303} = 0,772 \text{ atau } 77,2\%$$

$$\text{RECALL: } \frac{TP}{TP + FN} = \frac{234}{234 + 45} = \frac{234}{279} = 0,839 \text{ atau } 83,9\%$$

$$F - \text{Measure: } 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} = 2 \frac{0,772 \times 0,839}{0,772 + 0,839} = 0,804 \text{ atau } 80,4\%$$

Nilai akurasi, presisi, *recall*, dan *f-measure* yang dihitung secara manual untuk setiap kelas memberikan gambaran tentang performa model klasifikasi yang digunakan. Hasilnya adalah sebagai berikut:

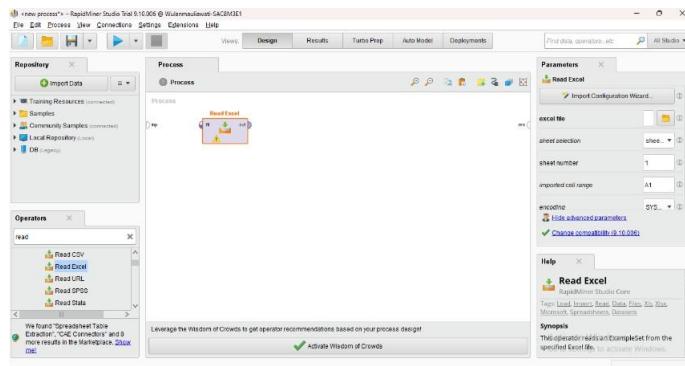
Tabel 5. Hasil

Kelas	Akurasi	Presisi	Recall	F-measure
Anemia hipokromik normositik	91.1%	77.2%	83.9%	80.4%
Anemia defisiensi besi	96.0%	90.4%	83.7%	87.0%
Anemia mikrositik	97.9%	62.1%	87.8%	72.8%
Leukemia	97.7%	46.8%	84.6%	60.3%
Sehat	94.5%	90.2%	89.0%	89.6%
Trombositopenia	98.4%	89.0%	83.3%	86.0%
Anemia normokromik normositik	94.9%	87.0%	88.3%	87.7%
Leukemia with Trombositopenia	99.5%	54.5%	85.7%	66.7%
Anemia makrositik	99.2%	66.7%	70.6%	68.6%

4.1.3 Perhitungan Algoritma K-Nearest Neighbor Menggunakan RapidMiner

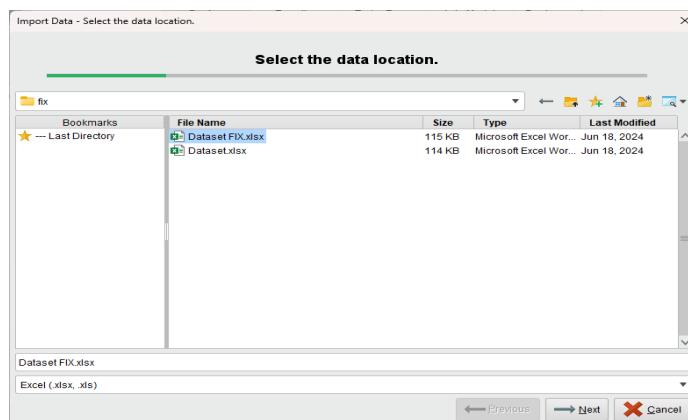
Pengujian yang dilakukan menggunakan *Rapidminer*. Masuk ke *RapidMiner Studio*.

- Impor* data gunakan operator "Read Excel" untuk mengimpor dataset.



Gambar 2. Import Read Excel

- Pilih dataset akses "Import Configuration Wizard" untuk memilih sel yang ingin diimporkan.



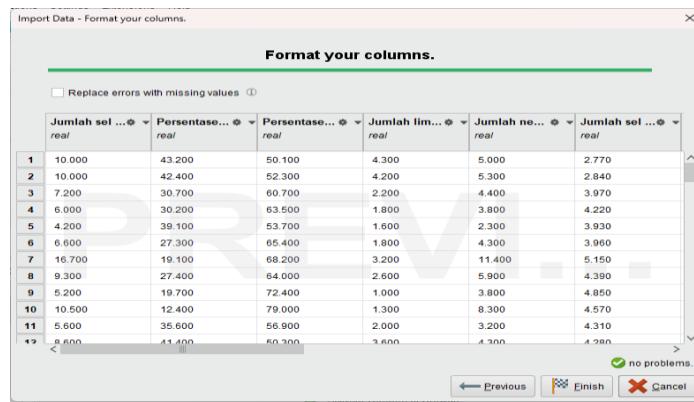
Gambar 3. Import Configuration Wizard "Dataset"

- Sesuaikan format data dengan kebutuhan yang akan digunakan.

Import Data - Select the cells to import.											
Select the cells to import.											
Sheet	Sheet1	Cell range:	A:O	Select All	<input checked="" type="checkbox"/> Define header row:	1					
1	JSDP	PL	PN	JL	JN	JSM	HB	HTK	VRSM	RHPS	
2	10.000	43.2	50.1	4.3	5.000	2.77	7.3	24.2	87.7	26.3	
3	10.000	42.4	52.3	4.2	5.3	2.84	7.3	25.000	88.2	25.7	
4	7.2	30.7	60.7	2.2	4.4	3.97	9.000	30.5	77.000	22.6	
5	6.000	30.2	63.5	1.8	3.8	4.22	3.8	32.8	77.9	23.2	
6	4.2	39.1	53.7	1.6	2.3	3.93	0.4	316.000	80.6	23.9	
7	6.6	27.3	65.4	1.8	4.3	3.96	8.8	29.7	75.2	22.2	
8	16.7	19.1	68.2	3.2	11.4	5.15	14.2	44.8	87.1	27.5	
9	9.3	27.4	64.000	2.6	5.9	4.39	12.000	37.9	86.4	27.3	
10	5.2	19.7	72.4	1.000	3.8	4.85	13.2	41.000	84.7	27.2	
11	10.5	12.4	79.000	1.3	8.3	4.57	12.4	38.9	85.3	27.1	
12	5.6	35.6	56.9	2.000	3.2	4.31	11.000	36.3	84.3	25.5	
13	8.6	41.4	60.3	3.6	4.3	4.28	11.000	35.000	81.9	25.7	

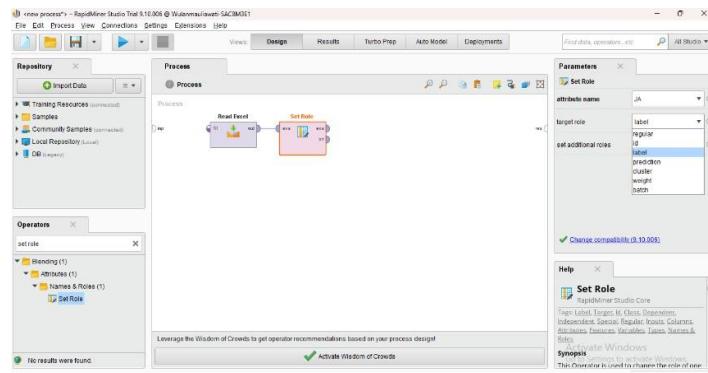
- Setelah data sesuai lalu klik finish.

Gambar 4. Pilih Sel Yang Ingin di Import



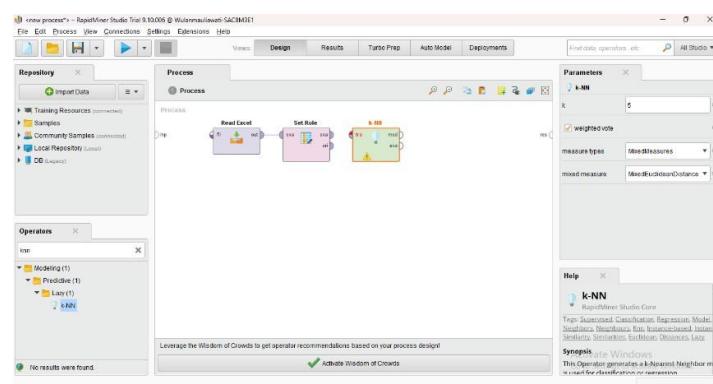
Gambar 5. Format Your Columns

- e. Tambahkan Set Role untuk setiap kolom data.



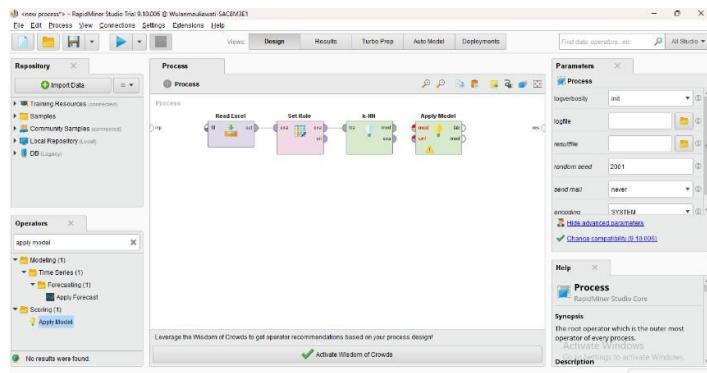
Gambar 6. Set Role

- f. Tambahkan operator K-Nearest Neighbor (KNN) untuk klasifikasi.

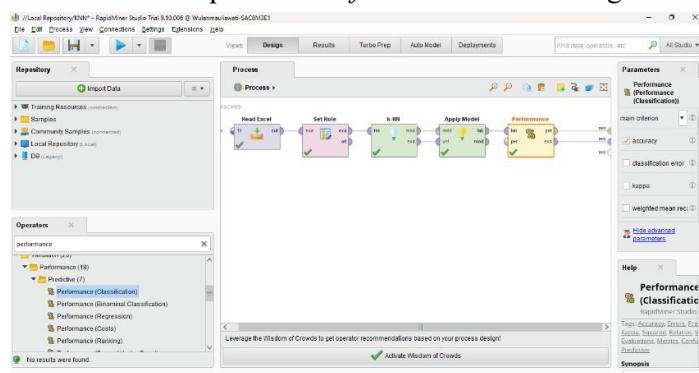


Gambar 7. K-NN

- g. Gunakan operator "Apply Model" untuk menerapkan model yang telah dibuat.

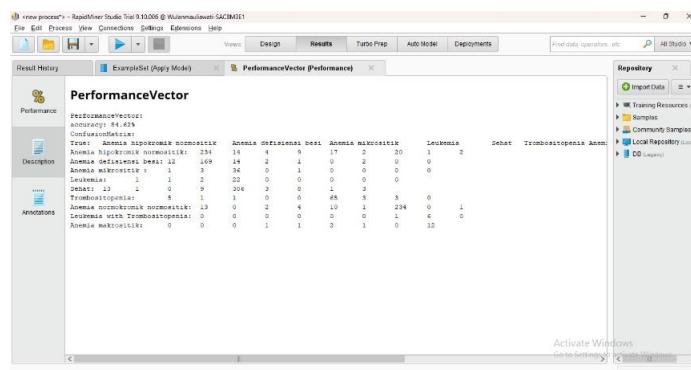
Gambar 8. *Apply Model*

h. Akhiri dengan menambahkan operator "*Performance*" untuk mengevaluasi hasil model.

Gambar 9. *Performance*

4.1.4 Validasi Hasil K-Nearest Neighbor

Setelah menjalankan proses evaluasi model menggunakan algoritma K-Nearest Neighbor (K-NN) di *RapidMiner*, hasil kinerja model ditunjukkan oleh *Performance Vector*. Dari hasil tersebut, model memiliki akurasi sebesar 84.62% yang dapat dilihat pada gambar VI.9.

Gambar 10. Hasil *Performance Vector*

4.1.5 Validasi Hasil K-Nearest Neighbor & Naive Bayes

Penelitian ini menerapkan K-Nearest Neighbor (K-NN) untuk mengklasifikasikan anemia dan kondisi kesehatan, dan dibandingkan dengan Naive Bayes. K-NN menunjukkan akurasi 84,62%, jauh lebih baik dari pada Naive Bayes yang hanya 52,54%. K-NN mengklasifikasikan lebih banyak kasus dengan benar dan memiliki kesalahan klasifikasi lebih

sedikit. K-NN terbukti lebih unggul dalam akurasi dan distribusi klasifikasi, menekankan pentingnya pemilihan algoritma yang tepat untuk data medis.

Tabel 6. Perbandingan K-NN & Naïve Bayes

Metode	Akurasi	Anemia hipokromik normositik	Anemia defisiensi besi	Anemia mikrositik	Leukemia	Sehat	Trombositopenia	Anemia normokromik normositik	Leukemia with Trombositopenia	Anemia makrositik
Naive Bayes	52,54%	3	148	5	8	275	28	80	2	10
KNN	84,62%	234	169	36	22	306	65	234	6	12

4.2 Pembahasan

Di bab ini, hasil penelitian menggunakan algoritma K-Nearest Neighbor (KNN) dengan perangkat *RapidMiner* untuk menganalisis dan mengklasifikasikan data. Data terdiri dari 1.281 entri dengan 15 atribut dan 1 atribut khusus yang mencakup 9 kelas anemia. Pengolahan data melibatkan perhitungan jarak *Euclidean*, seperti antara baris 1 dan baris 2 yang sebesar 2,61. Evaluasi model dilakukan dengan *Confusion Matrix* untuk mengukur akurasi, presisi, *recall*, dan *F-Measure*. Misalnya, untuk Anemia hipokromik normositik, diperoleh akurasi 91,1%, presisi 77,2%, *recall* 83,9%, dan *F-Measure* 80,4%. Hasil ini menunjukkan bahwa KNN efektif dalam klasifikasi anemia berdasarkan parameter CBC, memberikan kontribusi penting dalam diagnosis medis berbasis data.

5. PENUTUP

Penggunaan algoritma K-Nearest Neighbor (KNN) dalam klasifikasi jenis anemia menunjukkan hasil yang menjanjikan dengan akurasi keseluruhan 84,62%. Akurasi spesifik untuk berbagai jenis anemia adalah sebagai berikut: anemia hipokromik normositik 91,1% (presisi 77,2%, *recall* 83,9%), anemia defisiensi besi 96,0% (presisi 90,4%, *recall* 83,7%), anemia mikrositik 97,9% (presisi 62,1%, *recall* 87,8%), leukemia 97,7% (presisi 46,8%, *recall* 84,6%), kondisi sehat 94,5% (presisi 90,2%, *recall* 89,0%), dan trombositopenia 98,4% (presisi 89,0%, *recall* 83,3%). Hasil ini menunjukkan bahwa KNN efektif dalam mengklasifikasikan jenis anemia dengan akurasi tinggi, menjadikannya alat yang andal untuk diagnosis dan penanganan anemia.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah berperan dalam pembuatan jurnal implementasi algoritma K-Nearest Neighbor (KNN) dalam kategori penyakit anemia. Kerjasama yang baik dari semua pihak telah memastikan bahwa proses ini berjalan lancar sesuai dengan harapan. Khususnya, saya ingin menyampaikan rasa terima kasih yang mendalam kepada Ibu Ina Naiyah, M.kom selaku dosen pembimbing. Tanpa bimbingan dan arahan beliau, penulisan jurnal ini tidak akan mencapai standar yang dibutuhkan.

DAFTAR PUSTAKA

- [1] Y. Aulya, J. A. Siauta, and Y. Nizmadilla, “ANALISIS ANEMIA PADA REMAJA PUTRI,” *Jurnal Penelitian Perawat Profesional*, pp. 1377–1386, 2022, [Online]. Available: <http://jurnal.globalhealthsciencegroup.com/index.php/JPPP>
- [2] K. D. P. Sutanegara and D. Rahmadhona, “Anemia Aplastik: dari Awitan hingga Tatalaksana,” *Unram Medical Journal*, vol. 11, pp. 1094–1099, Feb. 2022, doi: 10.29303/jku.v11i3.768.
- [3] I. Lestari, M. Akbar, and B. Intan, “Perbandingan Algoritma Machine Learning Untuk klasifikasi Amenorrhea,” *Journal of Computer and Information Systems Ampera*, vol. 4,

- no. 1, pp. 32–43, Jan. 2023, doi: 10.51519/journalcisa.v4i1.371.
- [4] S. Widaningsih and S. Yusuf, “Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [5] F. Ramadhani, A. Satria, and I. P. Sari, “Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah,” *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 2, pp. 58–62, 2023.
- [6] S. Julaiha, M. Bettiza, and D. A. Purnamasari, “Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Klasifikasi Calon Penerima Bidikmisi,” *Student Online Journal (SOJ) UMRAH-Teknik*, vol. 2, no. 1, pp. 230–235, 2021.
- [7] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021.
- [8] Q. A'yuniyah and M. Reza, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru,” *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, 2023.
- [9] A. Algiffary and T. Sutabri, “Implementasi Machine Learning dengan Algoritma Naive Bayes Terhadap Sistem Informasi Pelayanan Pemberkasan Kepegawaian di BKPSDM Kota Palembang,” *Indonesian Journal of Computer Science*, 2023.
- [10] A. D. Sidik and A. Ansawarman, “Prediksi jumlah kendaraan bermotor menggunakan machine learning,” *Formosa Journal of Multidisciplinary Research*, vol. 1, no. 3, pp. 559–568, 2022.
- [11] R. W. Nasution, I. O. Suhada, I. G. Kirana, and I. P. Sari, “Penerapan data mining untuk pengelompokan minat konsumen terhadap pengguna jasa pengiriman pada pt. jalur nugraha ekakurir (jne) pematangsiantar,” *RESOLUSI Rekayasa Tek. Inform. dan Inf*, vol. 1, no. 4, pp. 274–281, 2021.
- [12] C. Zai, “Implementasi data mining sebagai pengolahan data,” *Jurnal Portal Data*, vol. 2, no. 3, 2022.
- [13] F. Marselina, I. Sofiyanti, A. R. Suryani, R. Pratiwi, T. Kariyani, and others, “Studi literatur: penyebab terjadinya anemia pada remaja putri: studi literatur: penyebab terjadinya anemia pada remaja putri,” in *Prosiding Seminar Nasional Dan CFP Kebidanan Universitas Ngudi Waluyo*, 2022, pp. 544–556.
- [14] M. Tamam, “Anemia of Chronic Disease: Key Points to Differentiate in Clinical Approach,” *Meet the Expert: Approach of Non Communicable Disease in Pediatric*, p. 16, 2021.
- [15] D. Pascalina, R. Widhiastono, and C. Juliane, “Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner,” *Technomedia J*, vol. 7, no. 3, pp. 293–302, 2022.
- [16] M. Murjani, “Prosedur Penelitian Kuantitatif,” *Cross-border*, vol. 5, no. 1, pp. 687–713, 2022.