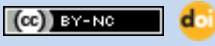


Analisis Performa Metode *K- Nearest Neighbor* (KNN) dan *Crossvalidation* pada Data Penyakit *Cardiovascular*

Inggrianti Pratiwi Putri^{a,1}

^a Program Studi Teknik Informatika, Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231, Indonesia

¹ 13020160264@umi.ac.id; ² Purnawansyah@umi.ac.id;

INFORMASI ARTIKEL	ABSTRAK
Diterima : 20 – 01 – 2021 Direvisi : 18 – 02 – 2021 Diterbitkan : 31 – 03 – 2021	Secara global, penyebab kematian nomor satu setiap tahunnya adalah penyakit <i>Cardiovascular</i> . Penyakit <i>cardiovascular</i> adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah, seperti Penyakit Jantung Koroner, Penyakit Gagal jantung atau Payah Jantung, <i>Hipertensi dan Stroke</i> . (Kemenkes RI, 2014) Tujuan dari penelitian ini adalah mengukur performa (akurasi, presisi, recall dan <i>f-measure</i>) metode knn dan <i>crossvalidation</i> pada <i>dataset cardiovascular</i> . dataset yang digunakan sebanyak 1000 record terdiri dari 11 atribut (<i>age, gender, height, dsb</i>) data pasien <i>cardiovascular</i> dan <i>non cardiovascular</i> , dataset tersebut diperoleh dari UCI <i>Machine Learning Repository</i> yang dikelola oleh Hungarian Institute of Cardiology Budapest: Andras Janosi, M.D., <i>University Hospital</i> , Zurich, Switzerland. Tahapan yang dilakukan yaitu: membagi rasio simulasi dataset (20:80, 50:50, 80:20), penerapan <i>crossvalidation</i> (<i>k-fold=10</i>) dan klasifikasi menggunakan metode K-NN (<i>k=2</i> hingga <i>K=900</i>), Hasil penelitian dari simulasi rasio dataset 50:50 memperoleh nilai akurasi 82%, presisi 82%, <i>recall</i> 82% dan <i>f-measure</i> 80% pada nilai <i>K=13</i> . Kemudian hasil penelitian dari simulasi rasio dataset 20:80 memperoleh nilai akurasi 87%, presisi 87%, <i>recall</i> 97% dan <i>f-measure</i> 92% pada nilai <i>K=3</i> . Dan hasil penelitian dari simulasi rasio dataset 80:20 memperoleh nilai akurasi 91%, presisi 92%, <i>recall</i> 60% dan <i>f-measure</i> 72% pada nilai <i>K=5</i> .
Kata Kunci: <i>K-Nearest Neighbor</i> <i>Crossvalidation</i> Akurasi Penyakit <i>Cardiovascular</i> .	

I. Pendahuluan

Cardiovascular merupakan organ manusia yang berperan dalam sistem peredaran darah. Penyakit *cardiovascular* adalah sebuah kondisi jantung tidak dapat melaksanakan tugasnya dengan baik. Secara global, penyebab kematian nomor satu setiap tahunnya adalah penyakit *cardiovascular*. [1] Penyakit *cardiovascular* adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah, seperti penyakit jantung koroner, penyakit gagal jantung atau payah jantung, *hipertensi and stroke*. Pada tahun 2013 diperkirakan sebanyak 12,3 juta kematian disebabkan oleh penyakit *cardiovascular*. [2] Lebih dari 3 juta kematian tersebut terjadi sebelum usia 60 tahun dan seharusnya dapat dicegah. Masyarakat berusia 20 hingga 40 menderita penyakit *cardiovascular* sebanyak 37% sedangkan yang berusia 40 hingga 60 sebanyak 71%. [3]

Penentuan kemiripan suatu data dapat di gun akan metode pengukuran jarak, salah satunya yaitu *K-Nearest Neighbor* (K-NN). K-NN merupakan metode pengukuran yang paling sering digunakan untuk menentukan kesamaan dua vektor [4][5][6].

Berdasarkan penelitian yang telah dilakukan sebelumnya, data yang digunakan 110 records pasien. 100 records digunakan sebagai data latih (*training data*) dan 10 records digunakan sebagai data uji (*testing data*). Untuk menentukan apakah seorang pasien terkena penyakit jantung digunakan 9 data terdekat atau *K= 9*. [7] Perhitungan kedekatan data *training* dengan kasus pada data *testing* 3 data diprediksikan masuk kedalam kelas “1” tetapi ternyata termasuk kedalam kelas “2”. Dari 100 data *training* dan 10 data testing dan menggunakan metode K-NN dengan nilai *K = 9* diperoleh tingkat akurasi sebesar 70%. [8]

Melanjutkan penelitian sebelumnya dengan menggunakan metode yang sama yaitu K-NN dengan jumlah data yang di tingkatkan sebesar 1000 data pasien serta melakukan simulasi nilai *K=3* hingga *K=9*. Hasil yang diperoleh dari penelitiannya yaitu nilai *K=6* memiliki nilai akurasi paling baik yaitu sebesar 85%, dan nilai presisinya 78%, *recall* 93%, dan *F-measure* 85%. Diakhir penelitiannya menyebutkan bahwa nilai akurasi

85% belum cukup baik maka di perlukan beberapa tindakan diantaranya menerapkan *crossvalidation* untuk simulasi data pada saat pengujian.[9]

Penelitian ini berupaya melanjutkan penelitian sebelumnya yaitu dengan menganalisis performa metode K-NN serta menerapkan *Crossvalidation*[1], [10], [11] pada data pasien *Cardiovascular*, harapan pada penelitian ini yaitu di perolehnya nilai performa yang baru dan dapat menjadi pembandingan dengan penelitian-penelitian sebelumnya.

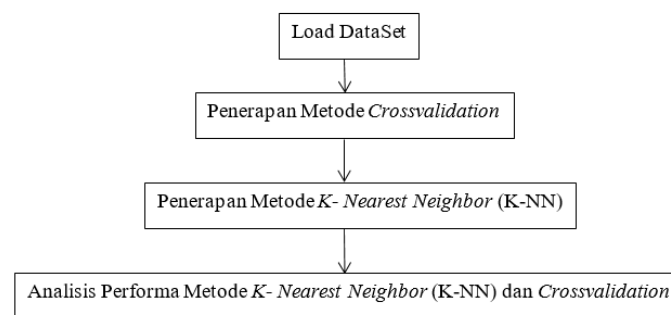
II. Metode

A. Perancangan Sistem

Pada penelitian ini secara garis besar langkah-langkahnya terdiri dari *load* dataset, penerapan *crossvalidation*, penerapan K-NN dan menghitung performa metode (akurasi, *preisis*, *recall*, dan *f-measure*) tersebut. memilah dataset menjadi data *training* dan data *testing* kemudian tahap persentasi data *training* kedalam implementasi K-NN pada data *testing* kemudian menghitung nilai akurasi, presisi, *recall*, dan *f-measure*.

B. Perancangan Proses

Adapun alur perancangan proses yang akan dilakukan ditunjukkan pada Gambar 1:



Gambar 1. Perancangan proses

1) Load Dataset Penyakit cardiovascular

Load dataset di bentuk menjadi 3 jenis yaitu dataset dengan rasio 50:50 (500 data *cardio* dan 500 data *non cardio*), rasio 20:80, (200 *cardio* dan 800 *noncardio*) dan rasio 80:20 (800 *cardio* dan 200 *non cardio*).

2) Penerapan metode crossvalidation

Crossvalidation berfungsi untuk membuat kombinasi dataset sehingga setiap data pada dataset berpeluang menjadi data *training* dan data *testing*, proses *crossvalidation* pada dataset dibagi menjadi 10 bagian dimana 900 dataset digunakan untuk data *training* dan 100 dataset di gunakan untuk data *testing*.

3) Metode K-Nearest Neighbor (K-NN)

Dalam proses pengklasifikasian ini untuk mengetahui apakah termasuk *cardio* atau *non cardio* dengan menggunakan rumus dari metode K-NN yang menggunakan *library scikit learn*.

4) Menghitung nilai Performa

Dalam Proses ini, kita melakukan perhitungan hasil nilai performa yang terdiri dari nilai akurasi, presisi, *recall* dan *f-measure* pada dataset penyakit *Cardiovascular*.

Performa merupakan bentuk tindakan, perbuatan, pekerjaan yang telah dicapai atau dilaksanakan, Kinerja dari klasifikasi dapat dievaluasi berdasarkan perhitungan performa nilai akurasi, presisi, *recall*, *f-measure*. [9]

1. Akurasi

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. rumus akurasi. Pada persamaan 2.

2. Presisi

Presisi didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut. rumus presisi ditunjukkan pada persamaan 3.

2. Recall

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Rumus *Recall* diuraikan pada persamaan 4.

3. *F-measure*

F-measure adalah *harmonic* mean antara nilai presisi dan recall, *F-measure* juga kadang disebut dengan nama *F-measure*. Rumus *F-measure* dijabarkan pada persamaan 5.

$$AKURASI = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots\dots\dots (1)$$

$$PRESISI = \frac{TP}{TP+FP} \dots\dots\dots (2)$$

$$RECALL = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

$$F - measure = 2 \frac{Presisi \times Recall}{Presisi+Recall} \dots\dots\dots (4)$$

Keterangan:

- TP : *True Positive*
- TN : *True Negative*
- FP : *False Positive*
- FN : *False Negative*

III. Hasil dan Pembahasan

A. *Implementasi Crossvalidation*

Penerapan *crossvalidation* dilakukan pada 3 dataset, dataset 1 dengan rasio 50:50 (500 *cardio* dan 500 *non cardio*), dataset 2 dengan rasio 20:80 (200 data *cardio* dan 800 *non cardio*), dataset 3 dengan rasio 80:20 (800 *cardio* dan 200 *non cardio*). Setiap dataset akan di bagi menjadi 10 bagian dimana setiap bagiannya terdiri dari 900 data *training* dan 100 data *testing*, setiap data akan bergiliran menjadi data *testing*.

B. *Implementasi Metode K-Nearest Neighbor (K-NN)*

Implementasi metode K-NN pada tahap ini merupakan contoh perhitungan manual mulai dari penetapan data *training* dan data *testing*, implementasi metode K-NN hingga perhitungan performa. Berikut ini 12 *sample data* yang akan diterapkan untuk perhitungan manual di tunjukkan pada Table 1 dan Table 2 menunjukkan 12 *sample data* yang akan di terapkan untuk perhitungan manual, 10 data *training* dan 2 data *testing*.

Tabel 1. Sampel Dataset (Data *Training*)

age	gender	Height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
16003	1	162	74.0	140	100	2	1	0	0	1	0
16039	2	180	90.0	140	90	2	2	0	0	0	0
19029	1	115	64.0	120	70	1	1	0	0	1	0
22821	2	168	80.0	160	100	1	1	0	0	1	0
20395	1	164	70.0	120	70	1	1	0	0	1	0
21473	2	174	90.0	140	80	1	1	1	1	0	1
21758	1	165	65.0	140	90	1	1	0	0	1	1
21166	2	162	59.0	140	90	2	1	1	0	1	1
23444	1	175	76.0	120	80	1	1	0	1	1	1
18958	1	149	62.0	180	110	2	1	0	0	1	1

Tabel 2. Sampel Data Set (Data *Testing*)

age	gender	Height	weight	ap_hi	ap_lo	Cholesterol	gluc	smoke	alco	active	cardio
17757	1	159	64.0	120	80	1	1	0	0	1	0
20994	1	156	72.0	140	100	2	1	0	0	1	1

Pada Tabel 1 dan Tabel 2 menunjukkan dataset yang digunakan dimana 10 data digunakan sebagai data *training* 2 sebagai data *testing*. Setelah ditentukan data *testing* tahap selanjutnya yaitu proses implementasi metode K-NN dengan menggunakan persamaan 1.

Pada Tabel 3 dan Tabel 4 menunjukkan hasil perhitungan data *testing* yang pertama. Dimana setiap satu data *testing* akan dihitung kesemua data *training*. Pada perhitungan dibawah ini merupakan hasil perhitungan dimana setiap atribut akan disandingkan pada atribut data *training* sesuai pada rumus *Eucludien* K-NN, kemudian di urutkan sesuai jumlah K yang telah ditentukan dimana hasil dari urutan tersebut dilihat *class*

yang paling dominan, dari hasil inilah yang menjadi acuan pengukuran performa dengan mencocokkan TP, TN, FP dan FN pada *Confusion matrix*[12]–[15].

Tabel 3. Atribut data *Testing*

<i>age</i>	<i>gender</i>	<i>Height</i>	<i>weight</i>	<i>ap_hi</i>	<i>ap_lo</i>	<i>Cholesterol</i>	<i>gluc</i>	<i>smoke</i>	<i>alco</i>	<i>active</i>	<i>cardio</i>
17757	1	159	64.0	120	80	1	1	0	0	1	0

Tabel 4. Perhitungan Data *Testing* 1

No	Hasil Data Testing 1	K=5	Urutan Hasil Perhitungan Jarak Terdekat	Hasil Klasifikasi	Actual
1.	1754,520212	4	1203,714	1	0
2.	1718,204055	3	1274,49	0	0
3.	1274,490094	2	1718,204	0	0
4.	5064,264354	9	1754,52	0	0
5.	2638,189543	5	2638,19	0	0
6.	3716,108337	7	3409,224	1	1
7.	4001,16758	8	3716,108	1	1
8.	3409,223965	6	4001,168	1	1
9.	15413,00734	10	5064,264	0	1
10.	1203,713839	1	15413,00734	1	1

Pada Tabel 5 dan Tabel 6 menunjukkan hasil perhitungan data *testing* yang kedua. Dimana setiap satu data testing akan dihitung kesemua data *training*. Pada perhitungan dibawah ini merupakan hasil perhitungan dimana setiap atribut akan disandingkan pada atribut data *training* sesuai pada rumus *Euclidian K-NN*, kemudian di urutkan sesuai jumlah K yang telah ditentukan dimana hasil dari urutan tersebut dilihat *class* yang paling dominan, dari hasil inilah yang menjadi acuan pengukuran performa dengan mencocokkan TP, TN, FP dan FN pada *Confusion matrix*.

Tabel 5. Atribut data *Testing*

<i>age</i>	<i>gender</i>	<i>Height</i>	<i>weight</i>	<i>ap_hi</i>	<i>ap_lo</i>	<i>Cholesterol</i>	<i>gluc</i>	<i>smoke</i>	<i>alco</i>	<i>active</i>	<i>cardio</i>
20994	1	156	72.0	140	100	2	1	0	0	1	1

Tabel 6. Perhitungan Data *Testing* 2

No	Hasil Data Testing 1	K=5	Urutan Hasil Perhitungan Jarak Terdekat	Hasil Klasifikasi	Actual
1.	4991,004	8	310,2934	1	0
2.	5635,563	9	946,0851	1	0
3.	4947,111	7	1208,563	1	0
4.	2373,111	5	1964,007	0	0
5.	1964,007	4	2373,111	0	0
6.	1208,563	3	3785,174	1	1
7.	946,0851	2	4947,111	0	1
8.	310,2934	1	4991,004	0	1
9.	19981	10	5634,563	0	1
10.	3785,174	6	19981	1	1

C. Pembahasan

Beberapa teknik pada penelitian ini menggunakan bahasa pemrograman python dan library scikitlearn mulai dari tahap load dataset *cardiovascular*, visualisasi data, *split data* implementasi simulasi *crossvalidation*, implementasi metode K-NN hingga menghitung nilai performa. Teknik *crossvalidation* di lakukan secara manual.

1) Pembahasan DataSet

Berdasarkan apa yang telah di utarakan sebelumnya bagian pengumpulan data bahwa dataset yang di gunakan pada penelitian ini berasal dari UCI *machine learning repository*. merupakan situs penyedia jasa *repository* dataset dimana dataset tersebut di kelolah oleh Hungarian Institute of Cardiology Budapest: Andras Janosi, M.D., *University Hospital, Zurich, Switzerland*. Yang mana data tersebut diunggah pada 1988 dan diunggah hingga tahun 2018. Pada Table 7 Menunjukkan simulasi pembagian data yang akan digunakan pada penelitian ini.

Tabel 7. Simulasi Pembagian Data Set

Dataset	Crossvalidation	Jumlah	Rasio	Split	
Dataset 1	I	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	II	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	III	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	IV	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	V	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	VI	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	VII	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	VIII	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	IX	1000	500 cardio	90% training	
			500 non cardio	10% testing	
	X	1000	500 cardio	90% training	
			500 non cardio	10% testing	
Dataset 2	I	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	II	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	III	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	IV	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	V	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	VI	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	VII	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	VIII	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	IX	1000	200 cardio	90% training	
			800 non cardio	10% testing	
	X	1000	200 cardio	90% training	
			800 non cardio	10% testing	
Dataset 3	I	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	II	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	III	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	IV	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	V	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	VI	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	VII	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	VIII	1000	800 cardio	90% training	
			200 non cardio	10% testing	
	IX	1000	800 cardio	90% training	
			200 non cardio	10% testing	
				800 cardio	90% training

	X	1000	800 cardio	10%testing
--	---	------	------------	------------

D. Pengujian Performa

Pengujian yang dilakukan pada penelitian ini diperuntukkan pada metode K-NN dan *crossvalidation*. Yang mana pengujian performa terdiri dari perhitungan nilai akurasi, presisi, *recall*, *f-measure*[5], [16], [17]. Hasil pengujian di tunjukkan dalam bentuk confusion matrix. Pada Tabel 5.1 menunjukkan nilai *confusion matriks* untuk nilai K= 57 pada simulasi jumlah *dataset cardio* dan *non cardio* dengan rasio 50:50.

Tabel 8. *Confusion matriks* rasio 50:50 Nilai K=57, Cv I

	Actual Terdiagnosa Cardiovascular	Actual Tidak Terdiagnosa Cardiovascular
Classified Terdiagnosa Cardiovascular	TP=43	FN=7
Classified Tidak Terdiagnosa Cardiovascular	FP=16	TN=34

Setelah di peroleh nilai TP, FP, FN dan TN maka nilai akurasi, presisi, *recall* dan *f-measure* dapat di hitung seperti yang tunjukkan berikut :

$$AKURASI = \frac{(43 + 34)}{(43 + 34 + 16 + 7)} = 0,77 = 77\%$$

$$PRESISI = \frac{43}{43 + 16} = 0,72 = 72\%$$

$$RECALL = \frac{43}{43 + 7} = 0,86 = 86\%$$

$$F - measure = 2 \frac{0,72 \times 0,86}{0,72 + 0,86} = 0,78 = 78\%$$

Pada Tabel 9. Menunjukkan nilai *confusion matriks* untuk nilai K= 3 pada simulasi jumlah dataset *cardio* dan *non cardio* dengan rasio 20:80.

Tabel 9. *Confusion matriks* rasio 20:80

	Actual Terdiagnosa Cardiovascular	Actual Tidak Terdiagnosa Cardiovascular
Classified Terdiagnosa Cardiovascular	TP=78	FN=2
Classified Tidak Terdiagnosa Cardiovascular	FP=11	TN=9

Setelah di peroleh nilai TP, FP, FN dan TN maka nilai akurasi, presisi, *recall* dan *f-measure* dapat di hitung seperti yang tunjukkan berikut.

$$AKURASI = \frac{(78 + 9)}{(78 + 9 + 11 + 2)} = 0,87 = 87\%$$

$$PRESISI = \frac{78}{78 + 11} = 0,87 = 87\%$$

$$RECALL = \frac{78}{78 + 2} = 0,97 = 97\%$$

$$F - measure = 2 \frac{0,87 \times 0,97}{0,87 + 0,97} = 0,92 = 92\%$$

Pada Tabel 10. Menunjukkan nilai *confusion matriks* untuk nilai K= 7 pada simulasi jumlah dataset *cardio* dan *non cardio* dengan rasio 80:20

Tabel 10. *Confusion matriks* rasio 80:20 Nilai K=8 Cv. II

	Actual Terdiagnosa Cardiovascular	Actual Tidak Terdiagnosa Cardiovascular
Classified Terdiagnosa Cardiovascular	TP=7	FN=13
Classified Tidak Terdiagnosa Cardiovascular	FP=4	TN=76

Setelah di peroleh nilai TP, FP, FN dan TN maka nilai akurasi, presisi, *recall* dan *f-measure* dapat di hitung seperti yang tunjukkan berikut.

$$AKURASI = \frac{(7 + 76)}{(7 + 76 + 4 + 13)} = 0,83 = 83\%$$

$$PRESISI = \frac{7}{7 + 4} = 0,63 = 63\%$$

$$RECALL = \frac{7}{7 + 13} = 0,35 = 35\%$$

$$F - measure = 2 \frac{0,63 \times 0,35}{0,63 + 0,35} = 0,45 = 45\%$$

Kesimpulan Pengujian dilakukan setelah proses metode K-NN dan nilai performa setiap dataset rasio 50:50 (500 dataset penyakit cardio dan 500 dataset penyakit *non cardio*), 20:80(200 dataset penyakit *cardio* dan 800 dataset penyakit *non cardio*) dan 80:20 (800 dataset penyakit *cardio* dan 200 dataset penyakit *non cardio*), telah dikumpulkan. Kesimpulan yang dihasilkan berupa nilai performa yang terdiri dari akurasi, presisi, *recall* dan *f-measure*. *Subpoin* berikut merepresentasikan keseluruhan hasil pengujian disertai dengan perbandingan hasil pengujian penelitian ini dengan penelitian sebelumnya.

IV. Kesimpulan

Berdasarkan hasil penelitian ini maka penulis dapat menarik kesimpulan bahwa Dataset1 (dataset 50:50) di peroleh nilai performa paling baik pada nilai akurasi sebesar 82%, presisi 82%, *recall* 82% dan *f-measure* 82%, pada K=13. Dataset2 (dataset 20:80) di peroleh nilai performa paling baik pada nilai akurasi sebesar 87%, presisi 87%, *recall* 97%, dan *f-measure* 92%, pada K=3. Dataset3 (dataset 80:20) diperoleh nilai performa paling baik pada nilai akurasi sebesar 91%, presisi 92%, *recall* 60% dan *f-measure* 72%, pada K=5.

Berdasarkan seluruh nilai yang telah didapatkan serta hasil perbandingan dengan penelitian sebelumnya, nilai performa yang diperoleh pada penelitian ini lebih baik di banding dengan nilai performa penelitian sebelumnya, serta menunjukkan bahwa nilai performa yang diperoleh dapat lebih baik karena dilakukan berbagai simulasi rasio data, penerapan crossvalidation serta pengujian keseluruhan nilai K pada K-NN.

Daftar Pustaka

- [1] H. Azis, "Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Indones. J. Data Sci.*, pp. 1–4, 2018, doi: <https://doi.org/10.33096/ijodas.v1i1.3>.
- [2] S. H. A. Aini, Y. A. Sari, and A. Arwan, "Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, 2018.
- [3] K. RI, "Situasi Kesehatan Jantung," *Pus. data dan Inf. Kementeri. Kesehat. RI*, 2014.
- [4] M. M. Baharuddin, T. Hasanuddin, and H. Azis, "Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, pp. 269–274, 2019.
- [5] H. Azis, F. T. Admojo, and E. Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah," *Techno.Com*, vol. 19, no. 3, 2020.
- [6] F. T. Admojo and Ahsanawati, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020.
- [7] M. Lestari, "Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung," *Fakt. Exacta*, vol. 7, no. September 2010, pp. 366–371, 2014.
- [8] H. Leidiana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, 2013.
- [9] C. A. U. Hassan, M. S. Khan, and M. A. Shah, "Comparison of Machine Learning Algorithms in Data classification," *Int. Conf. Autom. Comput.*, 2018.
- [10] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: [10.33096/ijodas.v1i3.20](https://doi.org/10.33096/ijodas.v1i3.20).
- [11] L. B. C. Tanujayaa, B. Susanto, and A. Saragiha, "Perbandingan Metode Regresi Logistik dan

-
- Random Forest untuk Klasifikasi Fitur Mode Audio Spotify,” *Indones. J. data Sci.*, vol. 1, no. 3, pp. 68–78, 2020.
- [12] D. Cahyanti, A. Rahmayani, and S. Ainy, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020.
- [13] Sugiarti and Mirnawati, “Implementasi Algoritma Government Standard (GOST) dalam Pengamanan File Dokumen,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 52–56, 2020.
- [14] Harlinda and Nasir, “Perancangan sistem pendukung keputusan dalam pengalokasian dana bantuan sosial di kabupaten pinrang dengan menggunakan metode AHP,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 44–51, 2020.
- [15] A. Prasetya Wibawa, W. Lestar, A. Bella Putra Utama, I. Tri Saputra, and Z. Nabila Izdihar, “Multilayer Perceptron untuk Prediksi Sessions pada Sebuah Website Journal Elektronik,” *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 57–67, 2020, doi: 10.33096/ijodas.v1i3.15.
- [16] M. M. Baharuddin, T. Hasanuddin, and H. Azis, “Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269–274, 2019.
- [17] L. Nurhayati and H. Azis, “Perancangan Sistem Pendukung Keputusan Untuk Proses Kenaikan Jabatan Struktural Pada Biro Kepegawaian,” *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 6–7, 2016.