

PENERAPAN ALGORITMA *NAÏVE BAYES* UNTUK MENDIAGNOSIS DAN *K-MEANS* UNTUK MENGELOMPOKKAN PENYAKIT GAGAL JANTUNG

Rakhmawati Hidayatur^{1*}, Nur Aeni Arinda²

¹Sistem Informasi STMIK MBP, ²Teknik Informatika STMIK MPB

*hidarahmawati@stmikmpb.ac.id

Abstrak

Penyakit gagal jantung menjadi salah satu penyakit yang memiliki angka kematian yang tinggi. Penyebab dari gagal jantung antara lain disfungsi miokard, endokard, perikardium, pembuluh darah besar, dan kelainan katup. Terjadinya kesalahan dalam mendiagnosis pasien menjadi permasalahan dalam penelitian ini. Adanya kemajuan teknologi dapat memberi kemudahan dalam segala aktivitas salah satunya dibidang kesehatan. Penelitian ini bertujuan untuk menerapkan algoritma naive bayes classifier untuk mendiagnosis benar tidaknya seorang pasien terkena gagal jantung dan k-means untuk mengelompokkan hasil dari diagnosis penyakit gagal jantung. Metode data mining yang digunakan pada penelitian ini adalah algoritma Naive Bayes untuk mendiagnosis dan algoritma K-Means untuk mengelompokkan penyakit gagal jantung. Data yang digunakan pada penelitian ini bersumber dari Rumah Sakit Umum Muhammadiyah Siti Aminah Bumiayu dengan jumlah 216 record dan 10 atribut diantaranya: gender, hypertension, heart_disease, ever_married, max_heart_rate, exang, cholestrol, status_smoking, slope, glucose. Hasil penelitian menunjukkan bahwa kinerja algoritma naive bayes menggunakan pengujian akurasi confusion matrix dengan nilai akurasi sebesar 80,22%. Berdasarkan hasil pengujian k-means diperoleh sebanyak 94 record yang terkena penyakit gagal jantung dari 216 record. 94 record tersebut diolah kembali menggunakan algoritma k-means menjadi 3 cluster, yaitu cluster rendah, sedang tinggi. Hasil dari klasterisasi yaitu cluster 0=27 record, cluster 1=50 record, cluster 2=17 record dengan mendapatkan validasi davies bouldin index (DBI) sebesar 1,450. Kesimpulannya yaitu penerapan algoritma naive bayes dan k-means dapat digunakan untuk diagnosis terkena penyakit gagal jantung dan mengklaster penyakit gagal jantung.

Kata Kunci: *Penyakit Gagal Jantung, Diagnosis, Data Mining, Naive Bayes, K-Means*

Abstract

Heart failure is one of the diseases that has a high mortality rate. Causes of heart failure include myocardial dysfunction, endocardium, pericardium, large blood vessels, and valve abnormalities. The occurrence of errors in diagnosing patients is a problem in this study. The existence of technological advances can make it easier in all activities, one of which is in the health sector. This study aims to apply the naïve Bayes classifier algorithm to correctly diagnose whether a patient has heart failure and k-means to group the results of the diagnosis of heart failure. The data mining method used in this research is the Naive Bayes algorithm to diagnose and the K-Means algorithm to classify heart failure diseases. The data used in this research came from the Muhammadiyah Siti Aminah Bumiayu General Hospital with a total of 216 records and 10 attributes including: gender, hypertension, heart_disease, ever_married, max_heart_rate, exang, cholestrol, status_smoking, slope, glucose. The results of the research show that the performance of the

Naive Bayes algorithm uses confusion matrix accuracy testing with an accuracy value of 80.22%. Based on the results of the k-means test, 94 records were found to be affected by heart failure out of 216 records. The 94 records were reprocessed using the k-means algorithm into 3 clusters, namely low and medium high clusters. The results of clustering are cluster 0=27 records, cluster 1=50 records, cluster 2=17 records with a validated Davies Bouldin Index (DBI) of 1.450. The conclusion is that the application of the Naive Bayes and K-Means algorithms can be used to diagnose heart failure and cluster heart failure.

Keywords: *Heart Failure, Diagnosis, Data Mining, Naive Bayes, K-Means*

1. Pendahuluan

Teknologi informasi mulai berkembang pada zaman media cetak dan penyiaran, dan terus berlanjut hingga suatu saat ketika masyarakat umum secara bertahap mulai mengetahui betapa pesatnya kemajuan teknologi informasi, seperti komputer, radio, televisi, dan lain-lain (Farhatun Nisaul Ahadiyah, 2023). IPTEK merupakan singkatan dari Teknologi Informasi dan Komunikasi, dimana teknologi dan pengetahuan semakin berkembang dan saling berhubungan serta mendukung perkembangan teknologi baru (Fitri & Haliza, 2021), salah satunya teknologi komputer. Pengguna teknologi komputer sebagai sarana dan prasarana serta sebagai penunjang dalam penggunaan sistem menghasilkan informasi kian membaik (Hanifah & Feizal, 2022), khususnya dalam bidang kesehatan. Pemanfaatan teknologi dalam bidang kesehatan sangat berpengaruh pada penyampaian pesan kesehatan khususnya pemberian informasi (Yuniar et al., 2022). Salah satu pemanfaatan teknologi dalam bidang kesehatan adalah diagnosis penyakit.

Diagnosis merupakan suatu proses menentukan kondisi atau penyakit apa yang dialami seseorang melalui penelitian serupa dan pemeriksaan gejala yang dialaminya (Iskandar, 2020). Manfaat diagnosis yaitu untuk menemukan kelemahan atau penyakit yang sedang di derita seseorang dan merupakan salah satu upaya untuk mencegah atau mengatasi penyebaran suatu penyakit (Sinurat, 2021). Salah satu kelemahan diagnosis adalah kemungkinan terjadinya *overlap* atau tumpang tindih antara diagnosis satu dengan diagnosis lainnya (Cherry Mustamu et al., 2023). Permasalahan yang sering terjadi yaitu kesalahan dalam

mendiagnosis penyakit pasien (Nasawida & Sari, 2022). Kesalahan diagnosis akan berdampak pada prosedur tindakan medis yang dapat merugikan pasien (Batista, 2024). Diagnosis penyakit pada penelitian ini diambil dari data Rumah Sakit Umum Muhammadiyah Siti Aminah Bumiayu (RSUMSA) tentang diagnosis penyakit gagal jantung.

Congestive Heart Failure (CHF) atau gagal jantung didefinisikan sebagai salah satu kumpulan gejala kompleks yang diakibatkan adanya gangguan pada proses kerja jantung, baik itu secara struktural maupun fungsional (Wardani et al., 2018). Data Riset Kesehatan Dasar (RisKesDas) Kementerian Kesehatan Indonesia pada tahun 2018 menyatakan bahwa penyakit gagal jantung di Indonesia semakin bertambah tiap tahunnya, dengan perkiraan sekitar 2.784.064 orang sehingga meningkat sebesar 0.13% dibandingkan dengan tahun 2013. Rumah Sakit Umum Muhammadiyah Siti Aminah Bumiayu menjadi salah satu contoh tingginya tingkat penyakit gagal jantung. Hasil wawancara yang dilakukan oleh peneliti dengan Kepala Rekam Medis (Maret 2024) menunjukkan bahwa ada 216 pasien penderita gagal jantung.

Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang besar (Arhami & Nasir, 2020). *Data mining* memiliki tujuh fungsi dasar yaitu *prediction* (prediksi), *sequencing* (sekuensi), *classification* (klasifikasi), *association* (asosiasi), *clustering* (pengelompokan), *forecasting*, *description* (deskripsi) (Mustika et al., 2021). Berdasarkan fungsi tersebut, penelitian ini menggunakan *classification* (klasifikasi) dan *clustering* (pengelompokan). Metode klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas

tertentu dari jumlah kelas yang tersedia (Utomo & Mesran, 2020). Salah satu algoritma untuk proses klasifikasi adalah algoritma *naïve bayes classifier*. Algoritma *naïve bayes* memiliki kelebihan diantaranya : cepat dan efisiensi ruang (Rini & Kunang, 2021). Tetapi algoritma *naïve bayes* memiliki kekurangan dimana probabilitasnya tidak akan bisa menghitung seberapa besar akurasi sebuah klasifikasi (Arifin & Ariesta, 2019). Selanjutnya, metode klusterisasi merupakan sebuah proses untuk mengelompokkan data ke dalam beberapa kelompok sehingga data dalam satu kluster memiliki tingkat kemiripan yang maksimum dan antar kluster memiliki kemiripan yang minimum (Sulastrri et al., 2021). *K-Means* adalah satu metode data *clustering* (Aulia, 2021). *K-Means* merupakan algoritma yang sederhana untuk diimplementasikan, memiliki kinerja yang relatif cepat, mudah beradaptasi dan umum digunakan (Hasibuan Budiansyah et al., 2022). Keunggulan dari algoritma *k-means* yaitu memiliki waktu komputasi yang relatif cepat (Ramadhani et al., 2022).

Berdasarkan uraian di atas, metode *naïve bayes* dan *k-means* dapat digunakan untuk membantu mendiagnosis penyakit gagal jantung, baik menggunakan metode *classification* maupun *clustering*. Oleh karena itu, penelitian ini akan menerapkan metode *naïve bayes classifier* untuk mendiagnosis benar tidaknya seorang pasien menderita penyakit gagal jantung, selanjutnya pasien yang terkena gagal jantung akan di input ke *k-means* untuk mengetahui tingkat rendah, sedang dan tinggi penyakit tersebut.

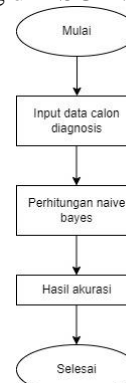
2. Metodologi Penelitian

Data bersumber dari data private yang diambil di Rumah Sakit Umum Muhammadiyah Siti Aminah Bumiayu. Pengambilan dataset dilakukan dari tanggal 13 Maret – 03 Mei 2024 dengan jumlah data sebanyak 216.

2.1 Eksperimen Algoritma Yang Digunakan

Penelitian ini mengusulkan algoritma *Naïve Bayes* dan *K-Means* sebagai algoritma

yang diusulkan untuk mrngklasifikasi penyakit gagal jantung di RSUD Bumiayu.



Gambar. 1. Eksperimen Naïve Bayes
Langkah-langkah eksperimen algoritma *naïve bayes* sebagai berikut:

1) Input Data

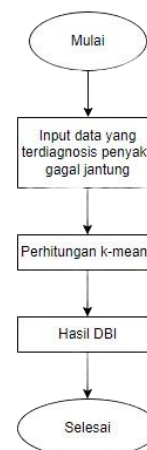
Input data merupakan langkah pertama dalam menerapkan algoritma *naïve bayes* dengan memasukkan *dataset* yang telah di dapatkan dari pengumpulan data.

2) Perhitungan Naïve Bayes

Perhitungan *naïve bayes* dengan menghitung nilai probabilitas setiap atribut.

3) Hasil Akurasi

Hasil akurasi dengan uji akurasi menggunakan *10-fold cross validation*.



Gambar 2. Eksperimen K-Means

1) Input Data

Input data merupakan langkah pertama dalam menerapkan algoritma *k-means* dengan memasukkan *dataset* terdiagnosis terkena gagal jantung yang telah didapatkan dari proses perhitungan *naïve bayes*.

2) Perhitungan *K-Means*

Menentukan jumlah *cluster* dan menghitung jarak dengan pusat *cluster awal*. Kemudian mengelompokkan hasil jarak terdekat antara data dengan pusat *cluster awal*. *Dataset* record yang sudah dihasilkan kemudia akan dihitung kembali sesuai dengan *cluster* untuk menghasilkan nilai *centroid* baru.

3) Hasil Akurasi

Hasil akurasi metode *k-means* menggunakan uji validasi *davies building index* (DBI) melalui aplikasi *rapidMiner*.

3. Hasil

Eksperimen yang dilakukan pada penelitian ini terdiri atas penerapan dan pengujian algoritma *naive bayes* dan *k-means* dalam mendiagnosis terkena penyakit gagal jantung. Pengujian algoritma *naive bayes* berupa uji akurasi menggunakan *confusion matrix* sedangkan *algoritma k-means* pengujian dilakukan menggunakan *10-fold cross validation*. Penerapan dan pengujian algoritma *naive bayes* dan *k-means* menggunakan 216 *record* dan 10 atribut dalam mendiagnosis penyakit gagal jantung.

3.1 Algoritma *Naive Bayes*

a. Perhitungan *Naive Bayes*

1) *Input Data*

Data penyakit gagal jantung dapat di lihat pada Tabel 1.

Tabel 1. Data Penyakit Gagal Jantung

No.	id	gender	hyper	heart_disease	ever_married
1	5492	male	104	no	yes
2	8717	female	80	no	yes
3	9421	female	120	no	yes
4	18008	female	103	no	yes
5	5492	male	104	no	yes
...
212	182474	female	173	yes	yes
213	182832	female	190	yes	yes
214	183249	male	126	no	yes
215	183523	female	183	yes	yes
216	183881	female	121	no	yes

max_heart_rate	ex	chol	status_smoking	slo	glu	status
82	0	195	yes	1	119	CHF
100	0	123	no	1	130	No CHF
100	0	109	no	2	126	No CHF
103	1	120	no	2	87	No CHF
82	0	195	yes	1	119	CHF
...
116	0	137	no	1	108	No CHF
90	0	119	no	1	65	No CHF
76	1	141	yes	3	76	No CHF
80	0	122	no	3	66	No CHF
102	1	224	no	3	89	CHF

2) Perhitungan *Naive Bayes*

Tabel 2 menunjukan jumlah *value dataset* yang akan dilakukan proses perhitungan nilai probabilitas setiap atribut:

Tabel 2. Jumlah *Value Dataset*

Atribut	Value (Faktor)	Jumlah	Label		Jumlah Kasus
			CHF (94)	No CHF (122)	
<i>Gender</i>	<i>Male</i>	81	37	44	216
	<i>Female</i>	135	57	78	
<i>Hypertension</i>	<150/90 mmHg	95	40	55	216
	<140/90 mmHg	121	54	67	
<i>Heart_disease</i>	<i>Yes</i>	61	38	23	216
	<i>No</i>	155	56	99	
<i>Ever_married</i>	<i>Yes</i>	191	81	110	216
	<i>No</i>	25	13	12	
<i>Max_heart_rate</i>	<65 bpm	20	9	11	216
	66-75 bpm	29	12	17	
	76-85 bpm	53	33	20	
	>86	114	40	74	

Exang	bpm				216
	0	177	75	102	
Cholestrol	1	39	19	20	216
	≤80 mg/dl	2	0	2	
	80-115 mg/dl	54	10	44	
Status_smo king	>115 mg/dl	160	88	72	216
	Yes	80	38	42	
Slope	No	136	56	80	216
	1	132	58	74	
	2	45	18	27	
Glucosa	3	40	19	20	216
	<90 mg/dl	57	17	40	
	90-99 mg/dl	29	15	14	
	100-109 mg/dl	34	17	17	
	110-125 mg/dl	96	47	49	

Penentuan dan perhitungan nilai probabilitas.

$$P(\text{CHF}) = \frac{92}{216} = 0,4351$$

$$P(\text{No CHF}) = \frac{122}{216} = 0,5648$$

Berikut nilai probabilitas setiap atribut:

- Gender

$$P(\text{Male}|\text{CHF}) = \frac{37}{94} = 0,3936$$

$$P(\text{Female}|\text{CHF}) = \frac{57}{94} = 0,6063$$

$$P(\text{Male}|\text{No CHF}) = \frac{44}{122} = 0,3606$$

$$P(\text{Female}|\text{No CHF}) = \frac{78}{122} = 0,6393$$

- Hypertension

$$P(<150|\text{CHF}) = \frac{40}{94} = 0,4255$$

$$P(<140|\text{CHF}) = \frac{54}{94} = 0,5744$$

$$P(<150|\text{No CHF}) = \frac{55}{122} = 0,4508$$

$$P(<140|\text{No CHF}) = \frac{67}{122} = 0,5249$$

Tabel 3 merupakan tabel peluang algoritma *naive bayes* dari nilai probabilitas masing-masing faktor.

Tabel 3 Peluang Naive Bayes

Atribut	Value (Faktor)	Label	
		CHF	No CHF
Gender	Male	0,3936	0,3606

Hypertension	Female	0,6063	0,6393
	<150/90 mmHg	0,4255	0,4508
	<140/90 mmHg	0,5733	0,5249
Heart_disease	Yes	0,4042	0,1885
	No	0,5957	0,8114
Ever_married	Yes	0,8671	0,901
	No	0,1382	0,098
Max_heart_rate	<65 bpm	0,095	0,0901
	66-75 bpm	0,127	0,139
	76-85 bpm	0,3510	0,163
	>86 bpm	0,425	0,606
Exang	0	0,797	0,836
	1	0,202	0,163
Cholestrol	≤80 mg/dl	0	0,0163
	80-115 mg/dl	0,1063	0,360
	>115 mg/dl	0,936	0,590
Status_smoking	Yes	0,404	0,344
	No	0,595	0,655
Slope	1	0,617	0,602
	2	0,1914	0,221
	3	0,202	0,163
Glucose	<90 mg/dl	0,1808	0,327
	90-99 mg/dl	0,1595	0,114
	100-109 mg/dl	0,1808	0,139
	110-125 mg/dl	0,5	0,401

3) Akurasi

Akurasi *Cross validation* yang diterapkan menggunakan *10-fold* yaitu *10-fold cross validation* berupa *10 fold data training*. Perhitungan untuk menentukan tingkat akurasi menggunakan *confusion matrix* pada dataset seperti Tabel 4.

Tabel 4 Akurasi dengan *10-fold cross validation*

		Actual Class	
		True CHF	True No CHF
Prediction Class	Pred CHF	68	16
	Pred No CHF	26	106

$$\text{Akurasi} = \frac{68+106}{68+16+26+106} \times 100 = 80,69\%$$

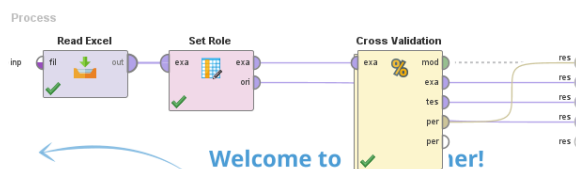
Nilai akurasi dari setiap *fold* seperti Tabel 4.5.

Tabel 4.5 Nilai Akurasi Setiap *Fold*

Fold	Dataset Diagnosis Terkena Penyakit Gagal Jantung									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

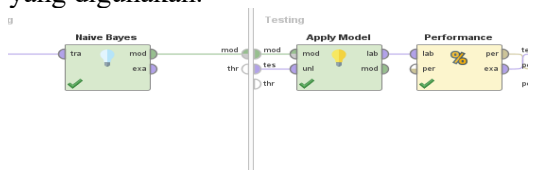
Fold 1 = 76,23%, fold 2 = 75,46%, fold 3 = 78,70%, fold 4 = 75,93%, fold 5 = 78,23%, fold 6 = 78,24%, fold 7 = 77,30%, fold 8 = 79,63%, fold 9 = 76,39%, fold 10 = 80,69%.

b. Hasil Rapidminer



Gambar 1. Eksperimen Naïve Bayes

Input data pasien penyakit gagal jantung menggunakan operator *read excel*. Data preprocessing menggunakan *set role* untuk memilih atribut yang akan digunakan sebagai label, parameter yang digunakan pada *set role attribute name=status*, target role=label. Model *cross validation* algoritma *naïve bayes* dimana operator *naïve bayes* untuk mendiagnosis penyakit gagal jantung dengan metode evaluasi berupa *10-fold cross validation* kemudian menggunakan operator *apply model* untuk menghasilkan *rule naïve bayes* dan operator *performance* untuk menghasilkan nilai akurasi dari kinerja model yang digunakan.



Gambar 2. Proses Performance

Hasil akurasi algoritma *naïve bayes* dalam diagnosis terkena penyakit gagal jantung menggunakan *rapidminer*.

accuracy: 80.69% +/- 9.35% (micro average: 80.56%)

	true CHF	true No CHF
pred. CHF	68	16
pred. No CHF	26	106
class recall	72.34%	86.89%

Gambar 3. Hasil Akurasi

3.2 Algoritma K-Means

a. Perhitungan K-Means

1) *Input Data*

Hasil diagnosis yang terkena penyakit gagal jantung sebanyak 94 *record* sedangkan yang terdiagnosis tidak terkena gagal jantung sebanyak 122 *record*. Berikut *dataset* yang terdiagnosis terkena penyakit gagal jantung sebanyak 94 *record* yang telah melalui proses klasifikasi.

Tabel 6. *Dataset* Diagnosis Gagal Jantung

No.	id	gender	hyper	heart_disease	ever_married
1	5492	male	104	no	yes
2	25282	female	160	yes	no
3	28238	female	179	yes	no
4	35072	male	158	yes	yes
5	41158	male	150	yes	yes
...
90	180515	male	174	yes	yes
91	191055	female	120	no	yes
92	181246	female	133	no	yes
93	181902	male	180	yes	yes
94	183881	female	121	no	yes

max_heart_rate	ex	chol	Status_smoking	slo	glu
82	0	195	yes	1	119
87	0	154	no	1	159
92	0	123	no	1	97
70	0	138	yes	3	117
100	0	110	yes	1	98
...
86	0	226	yes	1	101
185	1	177	no	2	105
90	0	190	no	3	281
84	0	180	yes	1	120
102	1	224	no	3	89

Hasil *record* yang sudah melalui

proses klasifikasi akan dilakukan proses perubahan type kategori ke dalam type numerik karena type numerik merupakan syarat untuk melakukan perhitungan dalam algoritma *k-means*.

Tabel 7. Dataset Numerik

No.	id	gender	hyper	heart_disease	ever_married
1	5492	1	2	2	1
2	25282	2	1	1	2
3	28238	2	1	1	2
4	35072	1	1	1	1
5	41158	1	1	1	1
...
89	180515	1	1	1	1
90	191055	2	2	2	1
91	181246	2	2	2	1
92	181902	1	1	1	1
93	183881	2	2	2	1

max_heart_rate	ex	chol	Status_smoking	slo	glu
3	1	3	1	1	4
4	1	3	2	1	4
4	1	3	2	1	2
2	1	3	1	3	4
4	1	2	1	1	2
...
4	1	3	1	1	3
4	2	3	2	2	3
4	1	3	2	3	4
3	1	3	1	1	4
4	2	3	2	3	1

2) Perhitungan *K-Means*

a) Penentuan *Cluster* Awal

Berdasarkan hasil penelitian dalam penentuan jumlah *cluster* untuk metode *k-means* berdasarkan uji validitas *davies bouldin index* (DBI) melalui perhitungan aplikasi *rapidminer* disajikan dalam Tabel 8.

Tabel 8. Jumlah *Cluster* dan Hasil DBI

No.	Jumlah nilai k	Hasil <i>Davies Bouldin Index</i> (DBI)
1.	1	1,483
2.	2	1,483
3.	3	1,450

Tabel 9. Pusat *Cluster* Awal

Centroid	gender	hyper	heart_disease	ever_married	max_heart_rate
C0	1	1	1	1	2
C1	1	1	1	1	1
C2	2	2	1	3	1
C3	1	1	1	1	4

ex	chol	Status_smoking	slo	glu
1	3	1	3	4
1	2	1	3	2
3	2	1	1	4
1	3	1	1	3

b) Perhitungan jarak awal dengan *cluster* awal

- 5492 (1,2,2,1,3,1,3,1,1,4)
C0 (1,1,1,1,1,1,3,1,1,1) =

$$\sqrt{(1-1)^2 + (2-1)^2 + (2-1)^2 + (1-1)^2 + (3-1)^2 + (1-1)^2} = 3,873$$

$$\sqrt{(1-1)^2 + (3-3)^2 + (1-1)^2 + (1-1)^2 + (4-1)^2} = 3,873$$

C1 (1,1,1,1,1,1,2,1,3,2) =

$$\sqrt{(1-1)^2 + (2-1)^2 + (2-1)^2 + (1-1)^2 + (3-1)^2} = 3,873$$

C2 (1,1,1,1,4,1,3,1,1,3) =

$$\sqrt{(1-1)^2 + (2-1)^2 + (2-1)^2 + (1-1)^2 + (3-4)^2} = 2$$

$$\sqrt{(1-1)^2 + (3-3)^2 + (1-1)^2 + (1-1)^2 + (4-3)^2} = 2$$

c) *Cluster* (Pengelompokkan) Data

Perhitungan hasil jarak akan dilakukan perbandingan dengan memilih jarak terdekat antara data dengan pusat *cluster* awal. Jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat awal *cluster* terdekat.

Tabel 10. Hasil Perhitungan Iterasi 1

No.	Id	C0	C1	C2	C3
1	5492	3,873	3,873	2,000	2,646
2	25282	4,583	4,583	2,000	3,317
3	28238	3,606	4,123	2,000	3,873
4	35072	3,742	2,449	3,000	0,000
5	41158	3,317	3,606	1,414	3,606
...
90	180515	3,606	3,873	0,000	3,000
91	191055	4,359	4,123	2,449	3,317
92	181246	5,099	4,243	3,000	2,828

93	181902	3,606	3,606	1,414	2,236
94	183881	4,243	4,000	3,606	4,243

Jarak Terdekat	Cluster
2,000	C2
2,000	C2
2,000	C2
0,000	C3
1,414	C2
...	...
0,000	C2
2,449	C2
2,828	C3
1,414	C2
3,606	C2

d) Penentuan Pusat Awal Cluster Baru

Tabel 11. Dataset Cluster Record

No.	Id	gender	hyper	heart_disease	ever_married	max_heart_rate
1	5492	1	2	2	1	3
2	25282	2	1	1	2	4
3	28238	2	1	1	2	4
4	35072	1	1	1	1	2
5	41158	1	1	1	1	4
...
90	180515	1	1	1	1	4
91	191055	2	2	2	1	4
92	181246	2	2	2	1	4
93	181902	1	1	1	1	3
94	183881	2	2	2	1	4

ex	chol	status_smoking	slop	glu	CLUSTER
1	3	1	1	4	C2
1	3	2	1	4	C2
1	3	2	1	2	C2
1	3	1	3	4	C3
1	2	1	1	2	C2
...
1	3	1	1	3	C2
2	3	2	2	3	C2
1	3	2	3	4	C3
1	3	1	1	4	C2
2	3	2	3	1	C2

Tabel diatas merupakan *dataset record* yang telah diklasterisasi. *Record* dari masing-masing atribut akan dihitung kembali sesuai dengan *cluster* grupnya sehingga menghasilkan nilai *centroid* baru.

Penentuan titik *centroid* baru:

$$\text{ClusterCenter C0 (Gender)} = \frac{2 + 2 + 2 + 1 + 2 + 2 + 2 + 1 + 1 + 1}{10} = \frac{16}{10} = 1,6$$

$$\text{ClusterCenter C0 (Hypertension)} = \frac{1 + 2 + 1 + 2 + 2 + 2 + 1 + 1 + 2 + 1}{10} = \frac{15}{10} = 1,5$$

$$\text{ClusterCenter C0 (Heart_Disease)} = \frac{2 + 2 + 1 + 2 + 2 + 2 + 1 + 2 + 2 + 1}{10} = \frac{17}{10} = 1,7$$

$$\text{ClusterCenter C0 (Ever_Merried)} = \frac{2 + 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{10} = \frac{12}{10} = 1,2$$

$$\text{ClusterCenter C0 (Max_Heart_Rate)} = \frac{3 + 1 + 3 + 1 + 3 + 3 + 3 + 3 + 3 + 1}{10} = \frac{24}{10} = 2,4$$

$$\text{ClusterCenter C0 (Ex)} = \frac{1 + 1 + 1 + 1 + 1 + 2 + 1 + 2 + 1 + 1}{10} = \frac{12}{10} = 1,2$$

$$\text{ClusterCenter C0 (Chol)} = \frac{3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3}{10} = \frac{30}{10} = 3$$

$$\text{ClusterCenter C0 (Smoking_Status)} = \frac{2 + 2 + 2 + 1 + 2 + 2 + 2 + 1 + 1 + 1}{10} = \frac{16}{10} = 1,6$$

$$\text{ClusterCenter C0 (Slope)} = \frac{1 + 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{10} = \frac{11}{10} = 1,1$$

$$\text{ClusterCenter C0 (Glu)} = \frac{1 + 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{10} = \frac{11}{10} = 1,1$$

Centroid	gender	hyper	heart_disease	ever_married	max_heart_rate
C0	1,6	1,5	1,7	1,2	2,4
C1	1,71	1,29	1,43	1,00	2,14
C2	1,55	1,45	1,55	1,18	3,54
C3	1,71	1,62	1,71	1,00	2,52

ex	chol	status_smoking	slop	glu	Rata-Rata
1,2	3	1,6	1,1	1,1	1,64
1,29	2,57	1,71	2,86	1,71	1,77
1,16	2,91	1,54	1,25	3,23	1,94
1,29	2,95	1,71	2,29	4,00	2,08

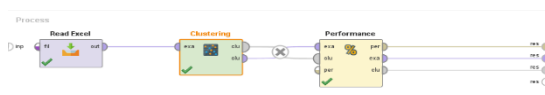
Tabel diatas merupakan hasil iterasi 1. Penentuan iterasi 2 caranya hampir sama, yaitu menghitung kembali anggota *cluster* dengan cara menentukan nilai jarak pada setiap objek dengan hasil *centroid* baru yang telah diperoleh. Berdasarkan hasil anggota *cluster* yang diperoleh berubah antara iterasi 1

dan 2, jadi iterasi dilanjutkan hingga menemukan batas iterasi maksimum. Iterasi maksimum terjadi ketika hasil anggota *cluster* yang diperoleh hasil *cluster* tidak berubah. Berdasarkan perhitungan yang telah dilakukan, iterasi maksimum terjadi pada iterasi ke-6 dimana hasil anggota *cluster* yang diperoleh tidak berubah antara iterasi 5 dan 6, jadi iterasi dihentikan. Berikut *centroid* baru dan hasil *cluster* pada iterasi ke-6.

Tabel 11. Hasil Iterasi 6

N o.	Id	gender	hyper	heart_disease	ever_married	max_heart_rate
1	5492	1	2	2	1	3
2	25282	2	1	1	2	4
3	28238	2	1	1	2	4
4	35072	1	1	1	1	2
5	41158	1	1	1	1	4
...
90	180515	1	1	1	1	4
91	191055	2	2	2	1	4
92	181246	2	2	2	1	4
93	181902	1	1	1	1	3
94	183881	2	2	2	1	4

b. Hasil Rapidminer



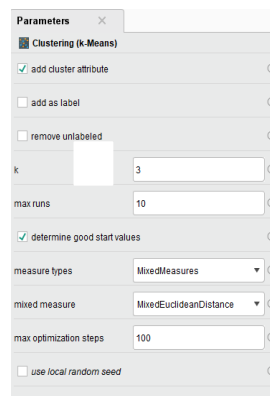
Gambar 4. Eksperimen K-Means

Input data pasien terdiagnosis penyakit gagal jantung menggunakan *read excel*. Pengelompokan algoritma *k-means* menggunakan operator *clustering*, penentuan parameter yang terdiri dari $k=3$, *measure types* = *MixedMeasures*, dan *mixed measure* = *MixedEuclideanDistance*. Hasil proses *k-means clustering* menggunakan 3 cluster model yakni cluster 0 = 27 record, cluster 1 = 50 record, cluster 2 = 17 record.

Cluster Model

```
Cluster 0: 27 items
Cluster 1: 50 items
Cluster 2: 17 items
Total number of items: 94
```

Gambar 4. Cluster Model



Gambar 5. Parameter *Clustering Centroid table* yang ada pada *rapidminer* merupakan tampilan berupa nilai untuk titik *centroid* dalam *cluster* pada sebuah tabel. Tampilan tersebut menunjukkan bahwa *cluster 3* merupakan *cluster* rendah, sedang dan *cluster 1* merupakan *cluster* tertinggi.

Attribute	cluster_0	cluster_1	cluster_2
gender	1.593	1.660	1.471
hyper	1.333	1.560	1.471
heart_disease	1.481	1.660	1.588
ever_married	1.148	1.140	1.059
max_heart_rate	3.074	3.600	1.588
ex	1.222	1.220	1.118
chol	2.889	2.920	2.882
status_smoking	1.556	1.660	1.471
slope	1.667	1.540	1.588
glu	1.444	3.780	3.529

Gambar 6. Centroid Table

Nilai *davies bouldin Index* (DBI) yang bekerja untuk mengevaluasi kinerja dari proses pengklasteran dimana evaluasi dari klasterisasi dari diagnosis terkena penyakit gagal jantung menghasilkan nilai sebesar 1,450

Davies Bouldin

Davies Bouldin: 1.450

Gambar 7. DBI

Hasil *eksperimen* algoritma *naive bayes* dengan pengujian *eksperimen* menggunakan *10-fold cross validation* dengan mengukur akurasi menggunakan *confusion matrix* menghasilkan nilai akurasi 80,69%. Hasil *eksperimen k-means* dengan menggunakan variasi jumlah *cluster* dengan mengukur kualitas *cluster* menggunakan *davies building index* menghasilkan 1,450.

4. Hasil

Penelitian ini menerapkan antara 2 algoritma data mining yaitu algoritma naïve bayes dan k-means terkena penyakit gagal jantung. Algoritma naïve bayes digunakan untuk mendiagnosis dan k-means digunakan untuk mengelompokkan penyakit gagal jantung. Pengujian akurasi menggunakan aplikasi rapidminer menggunakan model cross validation dan confusion matrix.

Pengujian kinerja algoritma naïve bayes menggunakan pengujian akurasi confusion matrix dengan nilai akurasi sebesar 80,69%. Berdasarkan hasil diagnosis diperoleh sebanyak 94 record yang terkena penyakit gagal jantung dari 216 record. 94 record tersebut diolah kembali menggunakan algoritma k-means menjadi 3 cluster, yaitu cluster rendah, sedang dan tinggi. Hasil klasterisasi tersebut cluster 0 = 27 record, cluster 1 = 50 record, cluster 2 = 17 record dengan mendapatkan validasi davies bouldin index (DBI) sebesar 1,450. Dengan demikian penerapan algoritma naïve bayes dan k-means dapat digunakan sebagai alternatif untuk diagnosis terkena penyakit gagal jantung.

5. Saran

Penelitian yang dilakukan dalam menerapkan algoritma *naïve bayes* dan *k-means* dalam diagnosis terkena penyakit gagal jantung telah berhasil. Perbaikan dan pengembangan untuk penelitian lanjutan adalah penelitian ini dapat dikembangkan dengan dibuatkannya sebuah aplikasi dalam mendiagnosis penyakit gagal jantung sekaligus mengkategorikan termasuk *cluster* rendah, sedang dan tinggi.

Daftar Pustaka

- Arhami, M., & Nasir, M. (2020). *Data Mining.Pdf*.
- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/jti.v13i1.97>
- Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. <https://doi.org/10.46576/djtechno.v1i1.964>
- Batista, G. T. (2024). Tinjauan Yuridis Tentang Pertanggungjawaban Kesalahan Diagnosis Dokter Pada Pelayanan Kesehatan. *Jurnal Hukum, Politik Dan Kekuasaan*, 4(2), 152–174. <https://doi.org/10.24167/jhpk.v4i2.10023>
- Cherry Mustamu, A., Alvian Fabonya, R., Mobalen, O., & Frisca Djamanmona, R. (2023). *Diagnosis 2.Pdf* (M. Nasrudin, Ed.).
- Farhatun Nisaul Ahadiyah. (2023). Perkembangan Teknologi Infomasi Terhadap Peningkatan Bisnis Online. *INTERDISIPLIN: Journal of Qualitative and Quantitative Research*, 1(1), 41–49. <https://doi.org/10.61166/interdisiplin.v1i1.5>
- Fitri, & Haliza, N. (2021). *Research & Learning in Faculty of Education Analisis Perkembangan Ilmu Pengetahuan dan Teknologi (Iptek) Dalam Pendidikan*. 3.
- Hanifah, A., & Feizal, M. (2022). ... Bangun Sistem Informasi Perkembangan Akademik Pada Pendidikan Anak Usia Dini (Paud) Berbasis Web Pada Satuan Paud Sejenis (Sps) *OKTAL: Jurnal Ilmu Komputer Dan ...*, 1(07), 997–1006.
- Hasibuan Budiansyah, R., Hafizah, H., & Mahyuni, R. (2022). Penerapan Data Mining Clustering Dengan Menggunakan Algoritma K-Means Pada Data Nasabah Kredit Bermasalah PT. BPR Milala. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 5(1), 7. <https://doi.org/10.53513/jsk.v5i1.4767>
- Iskandar, A. A. (2020). Diagnosa Penyakit Parasit pada Kucing Menggunakan Metode Certainty Factor (Studi Kasus : Puskewan Cibadak Kabupaten Sukabumi). *Jurnal Teknik Informatika Kaputama (JTik)*, 4(2), 126–134.
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Habis, I., Guntoro, Agnes Manuhutu, M., Ridwan, M., Hozairi, Khriana Wardhani, A., Alim, S., Romli, I., Religia, Y., Octafian, D. T., Utan Sufandi, U., & Ernawati, Ii. (2021). *data*

- mining 2.pdf (N. Rismawati, Ed.).
- Nasawida, M. D. P., & Sari, S. H. (2022). Penyelesaian Sengketa Medis Kesalahan Diagnosis. *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, 2(1), 10–27. <https://doi.org/10.53363/bureau.v2i1.12>
- Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 24–33. <https://doi.org/10.31849/digitalzone.v13i1.9292>
- Rini, O., & Kunang, S. O. (2021). Implementasi Data Mining Menggunakan Metode Naive Bayes Untuk Penentuan Penerima Bantuan Program Indonesia Pintar (Pip) (Studi Kasus : Sd Negeri 9 Air Kumbang). *Bina Darma Conference on ...*, 714–722.
- Sinurat, J. (2021). Jaringan Saraf Tiruan Diagnosa Penyakit Kanker Paru-Paru Menggunakan Metode Hebb Rule. *Bulletin of Information Technology (BIT)*, 2(1), 20–27.
- Sulastri, S., Usman, L., & Syafitri, U. D. (2021). K-prototypes Algorithm for Clustering Schools Based on The Student Admission Data in IPB University. *Indonesian Journal of Statistics and Its Applications*, 5(2), 228–242. <https://doi.org/10.29244/ijsa.v5i2p228-242>
- 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. <https://doi.org/10.30865/mib.v4i2.2080>
- Wardani, H. N., Sibuea, S., Lampung, U., Ilmu, B., Keluarga, K., Kedokteran, F., & Lampung, U. (2018). Penatalaksanaan Holistik Pada Wanita Usia 48 Tahun Dengan Vertigo Melalui Pendekatan Kedokteran Keluarga Holistic Management 48 Years Old Woman With Vertigo Through The Family Medicine Approach. 13(36), 1131–1141.
- Yuniar, M. C., Safila, M. I., Putra, M., Asyraf, M. H., Amelia, N. D., Kunto, D., & Patria, A. (2022). Pengembangan Teknologi dalam Bidang Kesehatan. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 18(2), 49–52.