



Analisis Perbandingan Metode *K-Nearest Neighbor* dan *Naive Bayes Classifier* pada Data Set Penyakit Jantung

Sahar^{a,1}

^a Program Studi Teknik Informatika, Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231, Indonesia
¹13020150075@umi.ac.id

INFORMASI ARTIKEL	ABSTRAK
Diterima : 31 – 07 – 2020 Direvisi : 12 – 09 – 2020 Diterbitkan : 31 – 12 – 2020	Di Indonesia telah terjadi pergeseran kejadian penyakit jantung dan pembuluh darah dari urutan ke-10 tahun 1980 menjadi urutan ke-8 tahun 1986. Sedangkan penyebab kematian tetap menduduki peringkat ke-3. Dalam proses pengklasifikasian ini untuk mengetahui apakah termasuk penyakit jantung atau non penyakit jantung dengan menggunakan rumus dari metode <i>K-Nearest Neighbor</i> dan <i>Naive Bayes Classifier</i> yang menggunakan <i>library scikit learn</i> . Dalam proses penelitian ini kita melakukan perhitungan hasil nilai performa yang terdiri dari akurasi, presisi, <i>recall</i> dan <i>f-measure</i> pada dataset penyakit jantung. Menggunakan metode klasifikasi yg memiliki hasil uji performa tertinggi/terbaik. Berdasarkan hasil pengujian, didapatkan tingkat akurasi pada metode <i>K-Nearest Neighbor</i> sebesar 67%, presisi 65%, <i>recall</i> 73%, dan <i>f-measure</i> 96% pada nilai K=250 dan metode jarak Manhattan, tingkat akurasi pada metode jarak <i>Euclidean</i> sebesar 65%, presisi 65%, <i>recall</i> 69%, dan <i>f-measure</i> 67% pada nilai K=250 sedangkan pada metode <i>Naive Bayes Classifier</i> tingkat akurasi yang didapatkan sebesar 58%, presisi 90%, <i>recall</i> 55%, dan <i>f-measure</i> 68%. Performa metode klasifikasi terbaik pada dataset Penyakit jantung yaitu metode KNN (K-0).
Kata Kunci: <i>Box plot</i> <i>Python</i> <i>Euclidean distance</i> <i>Manhattan Distance</i> <i>Machine Learning</i>	 

I. Pendahuluan

Di Indonesia telah terjadi pergeseran kejadian penyakit jantung dan pembuluh darah dari urutan ke-10 tahun 1980 menjadi urutan ke-8 tahun 1986. Sedangkan penyebab kematian tetap menduduki peringkat ke-3. Meski belum ada data epidemiologis pasti, angka kesakitan/kematiannya terlihat cenderung meningkat. Hasil Survei Kesehatan Nasional tahun 2001 menunjukkan tiga dari 1.000 penduduk Indonesia menderita penyakit jantung.[1]

Jantung merupakan organ yang penting dalam sistem tubuh manusia. Jantung berfungsi untuk memompakan darah yang mengandung oksigen dan nutrien ke seluruh tubuh. Jantung terdiri dari beberapa ruang yang dibatasi oleh beberapa katup, diantaranya adalah katup atrioventrikular dan semilunar. Katup atrio ventrikular terdiri atas katup bicuspid (mitral) dan katup tricuspid, yang terletak di antara atrium dan ventrikel, sedangkan katup semilunar terletak antara ventrikel dengan aorta dan arteri pulmonal.

Jantung manusia memiliki 4 katup yang memiliki fungsi yang berbeda antar satu dengan yang lain. Kompleks mitral merupakan salah satu komponen dari jantung yang memiliki peran sangat penting.[2]

Penyakit Jantung merupakan salah satu bentuk penyakit kardiovaskular yang menjadi penyebab kematian nomor satu di dunia. Penyakit jantung adalah suatu penyakit degeneratif yang berkaitan dengan gaya hidup, dan sosial ekonomi masyarakat. Penyakit ini merupakan masalah kesehatan utama di negara maju. Badan Kesehatan Dunia (WHO) mencatat lebih dari 7 juta orang meninggal akibat penyakit jantung di seluruh dunia pada tahun 2002. Angka ini diperkirakan meningkat hingga 11 juta orang pada tahun 2020.

Pengumpulan data dan informasi sebagai acuan dalam melakukan penelitian, diantaranya adalah studi pustaka dan dataset dari UCI *Machine Learning Repository*. [3] Data tersebut diunggah pada tahun 1988 dan terus diupdate hingga tahun 2018. Adapun atribut yang menjadi acuan terjadinya penyakit jantung antara lain AIDS, *age*, *gender*, *weight*, *ap hi* (tekanan darah systolic), *ap lo* (tekanan darah diastolic), *cholesterol*, *glucose*, *smoke*, *alcohol*, *cardio*.

Data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data.[4] Data mining akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data.[5] Salah satu metode yang sering digunakan dalam mengolah data mining adalah metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*. [6]

K-Nearest Neighbor (KNN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok K objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data *testing*.

Naive Bayes Classifier merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas.[6]

Berdasarkan uraian sebelumnya, maka pada penelitian ini dilakukan penerapan metode *K-Nearest Neighbor* dan *Naive Bayes Classifier* terhadap data yang dikumpulkan oleh *Hungarian Institute of Cardiology*. Budapest: Andras Janosi, M.D., University Hospital, Zurich, Switzerland; William Steinbrunn, M.D., University Hospital, Basel, Switzerland; Matthias Pfisterer, M.D. dan V.A. Medical Center, Long Beach and Cleveland Clinic Foundation; Robert Detrano, M.D., Ph.D.

II. Metode

A. Dataset

Dataset yang digunakan pada penelitian ini adalah dataset penyakit cardiovascular yang diperoleh dari UCI *Machine Learning Repository* merupakan sebuah situs penyedia jasa *repository dataset*, dimana dataset yang digunakan pada penelitian ini dikelola oleh *Hungarian Institute of Cardiology*. Budapest: Andras Janosi, M.D., University Hospital, Zurich, Switzerland; William Steinbrunn, M.D., University Hospital, Basel, Switzerland; Matthias Pfisterer, M.D. dan V.A. Medical Center, Long Beach and Cleveland Clinic Foundation; Robert Detrano, M.D., Ph.D. yang mana data tersebut di unggah pada tahun 1988 dan terus di update hingga tahun 2019. Dataset tersebut terdapat 7000 sampel data pasien penyakit jantung dan non penyakit jantung, data pasien tersebut terdiri dari 11 atribut yaitu age(umur), gender(jenis kelamin), height(tinggi badan), weight(berat badan), ap_hi(tekanan darah sistolik), ap_lo(tekanan darah diastolik), kolesterol (kolesterol), gluc (gula darah), smoke (status perokok), alco (status peminum), active (status aktif).

Pada tahap selanjutnya yaitu penerapan algoritma *K-Nearest Neighbor* dan *Naive Bayes Classifier* sebagai metode atau algoritma klasifikasi. Hasil dari proses klasifikasi yang akan dijadikan acuan untuk menghitung performa metode *K-Nearest Neighbor* dan Metode *Naive Bayes Classifier*.

B. Perancangan Sistem

Dalam penelitian ini secara garis besar langkah-langkahnya terdiri dari load dataset, memilah dataset menjadi data training dan data testing kemudian tahap persentasi data training kedalam implementasi KNN dan NBC pada data testing kemudian menghitung dan membandingkan nilai akurasi, presisi, recall, dan f- measure.

1) Perancangan Proses Data mining

Adapun perancangan proses data mining Antara lain:

- Deskripsi

Mengidentifikasi pola yang muncul secara berulang pada data dan mengubah pola tersebut menjadi aturan dan kriteria.

- Klasifikasi

Proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas.

- Clustering

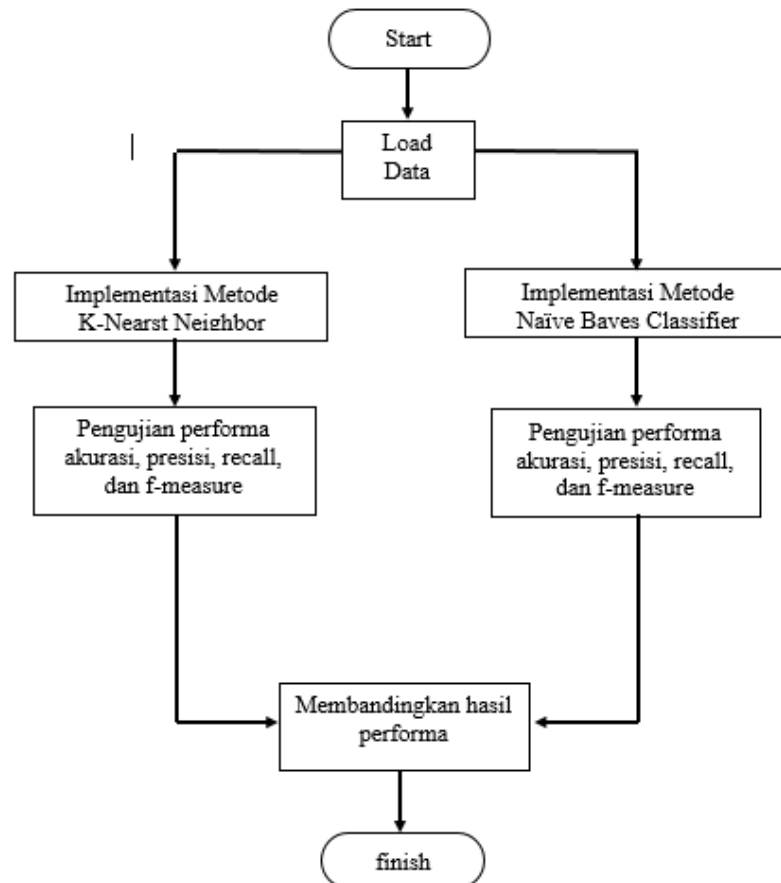
Pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama.

- Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu.

2) Perancangan Proses

Adapun alur perancangan proses yang akan ditunjukkan pada Gambar berikut.



Gambar 1. Alur Perancangan Proses

C. Metode Penelitian

1) Tahapan Penelitian

- Load dataset

Dataset yang digunakan pada penelitian ini berjumlah 70000 *record* dimana akan dibagi menjadi 5 macam berdasarkan pembagian jumlah data *training* dan data *testing* yaitu, 10% data *testing*, 20% data *testing*, 30% data *testing*, 40% data *testing*, 50% data *testing*.

- Penerapan metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*

Dalam proses pengklasifikasian ini untuk mengetahui apakah termasuk penyakit jantung atau non penyakit jantung dengan menggunakan rumus dari metode KNN dan *Naive Bayes Classifier* yang menggunakan library *scikit learn*.

- Menghitung performa metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*

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

- Membandingkan performa metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*

Pada tahap ini performa yang diukur pada kedua metode yaitu akurasi, presisi, *recall*, dan *f-measure* akan dibandingkan serta menarik kesimpulan performa metode yang terbaik

Metode berisi tahapan atau prosedur penelitian dan algoritma yang digunakan dalam penelitian, formula permasalahan yang diteliti dengan lebih rinci, serta perancangan sistem jika dibutuhkan.

III. Hasil dan Pembahasan

A. Implementasi metode *K-Nearest Neighbor*

Implementasi metode KNN pada tahap ini merupakan contoh perhitungan manual mulai dari penetapan data *training* dan data *testing*, implementasi metode KNN hingga perhitungan performa. Berikut ini 12 sample

data yang akan di terapkan untuk perhitungan manual di tunjukkan pada Table 1 dan 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*)

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
18393	2	168	62	110	80	1	1	0	0	1	0
20228	1	156	85	140	90	3	1	0	0	1	1
18857	1	165	64	130	70	3	1	0	0	0	1
17623	2	169	82	150	100	1	1	0	0	1	1
17474	1	156	56	100	60	1	1	0	0	0	0
21914	1	151	67	120	80	2	2	0	0	0	0
22584	2	178	95	130	90	3	3	0	0	1	1
19631	2	173	79	160	100	1	1	1	0	1	0
16131	2	176	71	110	70	1	1	1	1	1	0
21921	2	171	78	140	80	3	2	1	1	1	1

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

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
14	19834	1	164	68	110	60	1	1	0	0	0
15	22530	1	169	80	120	80	1	1	0	0	1

Pada Table 1 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 KNN dengan menggunakan persamaan 1.

Pada Tabel 3 dan 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 *Manhattan* KNN, 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*[7]–[9].

Tabel 3. Atribut data *Testing*

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
14	19834	1	164	68	110	60	1	1	0	0	0

Tabel 4. Perhitungan Data *Testing* 1

No	Hasil perhitungan jarak	Urutan terdekat	Label Actual
1	1441,158	4	0
2	396,7291	2	1
3	977,2666	3	1
4	2211,774	7	1
5	2360,065	8	0
6	2080,162	5	0
7	2750,406	9	1
8	213,3401	1	0
9	3703,035	10	0
10	2087,349	6	1

Pada Tabel 5 dan 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* KNN, 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* 2

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
15	22530	1	169	80	120	80	1	1	0	0	1

Tabel 6. Perhitungan Data *Testing 2*

No	Hasil Perhitungan Jarak	Urutan	
		Terdekat	Label Actual
1	4137,051	7	0
2	2302,152	4	1
3	3673,065	6	1
4	4907,133	8	1
5	5056,153	9	0
6	616,4025	3	0
7	58,57474	1	1
8	2899,348	5	0
9	6399,026	10	0
10	609,3414	2	1

B. Implementasi metode Naive Bayes Classifier

Implementasi metode *Naive Bayes Classifier* Pada tahap ini merupakan contoh perhitungan manual mulai dari penetapan data *training* dan data *testing*. implementasi metode *Naive Bayes Classifier* hingga perhitungan performa . berikut ini 12 sampel data yang diterapkan untuk perhitungan manual di tunjukkan Pada Tabel 7 dan 8 menunjukkan 12 sampel data untuk perhitungan manual, 10 data *trainig* dan 2 data *testing*.

Tabel 7. Sampel Dataset (Data *Training*)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
18393	2	168	62	110	80	1	1	0	0	1	0	0
20228	1	156	85	140	90	3	1	0	0	1	1	1
18857	1	165	64	130	70	3	1	0	0	0	0	1
17623	2	169	82	150	100	1	1	0	0	1	1	1
17474	1	156	56	100	60	1	1	0	0	0	0	0
21914	1	151	67	120	80	2	2	0	0	0	0	0
22584	2	178	95	130	90	3	3	0	0	1	1	1
19631	2	173	79	160	100	1	1	1	0	1	0	0
16131	2	176	71	110	70	1	1	1	1	1	0	0
21921	2	171	78	140	80	3	2	1	1	1	1	1

Tabel 8. Sampel Dataset (Data *Testing*)

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
14	19834	1	164	68	110	60	1	1	0	0	0
15	22530	1	169	80	120	80	1	1	0	0	1

Pada Tabel 7 dan 8 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 *Naive Bayes Classifier*.

Pada Tabel 9 menunjukkan hasil perhitungan *mean* (rata-rata) setiap atribut. Dimana data *training* yang ada didalam setiap atribut di jumlahkan setelah itu dibagi dengan jumlah dataset penyakit jantung atau non penyakit jantung dengan menggunakan rumus pada metode *Naive Bayes Classifier* pada persamaan 7.

Tabel 9. Perhitungan *Mean* (rata-rata)

Kelas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
Jantung	20243	2	168	81	138	86	3	2	0	0	1
Non Jantung	18709	2	165	67	120	78	1	1	0	0	1

Pada Tabel 10 menunjukkan hasil perhitungan standar deviasi setiap atribut. Dimana data *training* yang ada didalam setiap atribut dikurangkan dengan hasil perhitungan *mean* (rata-rata) dan dipangkat duakan. setelah itu dibagi dengan jumlah dataset penyakit jantung dan non penyakit jantung dengan menggunakan rumus pada metode *Naive Bayes Classifier* pada persamaan 8.

Table 11. Perhitungan Standar Deviasi

Kelas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
-------	----	----	----	----	----	----	----	----	----	-----	-----

Jantung	2066,49	0,54	8,10	11,30	8,36	11,40	0,89	0,89	0,44	0,44	0,44
Non Jantung	2201,89	0,54	10,8	8,74	23,45	14,83	0,44	0,44	0,54	0,44	0,54

Pada Tabel 11 menunjukkan hasil perhitungan normal distribusi dengan menggunakan data *testing* setiap atribut dengan menggunakan rumus pada metode *Naive Bayes Classifier* pada persamaan 9.

Tabel 12. Perhitungan *Testing*

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	Hasil Prediksi
14	19834	1	164	68	110	60	1	1	0	0	0	
Jantung	0,008608	0,912	0,1255	0,0624	0	0	0,085187625	0,336	0	0	0,1204	4,51
Non Jantung	0,007463	0,295	0,1208	0,1340	0,075	0,049	0,539924744	0,539	0	0	0,2959	1,28
											Maximal	1,28

C. Pembahasan

Beberapa teknik pada penelitian ini menggunakan bahasa pemrograman *Python* dan *library scikit learn*. mulai dari tahap *load* dataset penyakit jantung, visualisasi data implementasi metode *K-Nearest Neighbor* dan metode *Naive Bayes Classifier*, hingga mendapatkan nilai performa, dan menentukan perbandingan kedua metode yang lebih baik nilai performa.

1) Pembahasan dataset

Berdasarkan apa yang telah diutarakan pada bab 1 bagian pengumpulan data bahwa dataset yang digunakan pada penelitian ini berasal dari *UCI machine learning repository*. *UCI machine learning repository* merupakan sebuah situs penyedia jasa *repository* dataset. dimana dataset yang digunakan pada penelitian ini dikelola oleh *Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D., University Hospital, Zurich, Switzerland: William Steinbrunn, M.D., University Hospital, Basel, Switzerland Matthias Pfisterer, M.D. dan V.A. Medical Center, Long Beach and Cleveland Clinic Foundation Robert Detrano, M.D., Ph.D.* yang mana data tersebut di unggah pada tahun 1988 dan terus di update hingga tahun 2018.

D. Pengujian Performa

1) Performa *K-Nearest Neighbor*

Pengujian yang dilakukan pada penelitian ini di peruntukkan pada dua metode Antara lain metode KNN. Yang mana pengujian performa terdiri dari perhitungan nilai akurasi, presisi, recall, dan *f-measure*. Hasil pengujian di bagi menjadi 2 untuk pengukuran jarak manhattan dan pengukuran jarak *Euclidean* serta ditunjukkan dalam bentuk *Confusion Matrix*.

Tabel 13. Confusion matriks hasil pengujian metode KNN Manhattan

	Actual Terdiagnosa Penyakit Jantung	Actual Tidak Terdiagnosa Penyakit Jantung
Classified Terdiagnosa Penyakit Jantung	TP=2598	FN=922
Classified Tidak Terdiagnosa Penyakit Jantung	FP=1341	TN=2139

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

$$AKURASI = \frac{(2598 + 2139)}{(2598 + 1341 + 2139 + 922)} = 0,67 = 67\%$$

$$PRESISI = \frac{2598}{2598 + 1341} = 0,65 = 65\%$$

$$RECALL = \frac{2598}{2598 + 922} = 0,73 = 73\%$$

$$F - measure = 2 \frac{0,67 \times 0,65}{0,67 + 0,65} = 0,69 = 69\%$$

Tabel 14. Confusion matriks hasil pengujian metode KNN Euclidean

	Actual Terdiagnosa Penyakit Jantung	Actual Tidak Terdiagnosa Penyakit Jantung
Classified Terdiagnosa Penyakit Jantung	TP=2457	FN=1063
Classified Tidak Terdiagnosa Penyakit Jantung	FP=1319	TN=2161

Setelah di peroleh nilai TP, FP, FN dan TN maka nilai akurasi, presisi, recall dan *f-measure* dapat di hitung seperti yang tunjukkan berikut [10]–[14].

$$AKURASI = \frac{(2457 + 2161)}{(2457 + 1319 + 1063 + 2161)} = 0,65 = 65\%$$

$$PRESISI = \frac{2457}{2457 + 1319} = 0,65 = 65\%$$

$$RECALL = \frac{2457}{2457 + 1063} = 0,69 = 69\%$$

$$F - measure = 2 \frac{0,65 \times 0,69}{0,65 + 0,69} = 0,67 = 67\%$$

2) Performa Naïve Bayes Classifier

Pengujian yang dilakukan pada penelitian ini diperuntukkan pada dua metode antara lain metode *Naïve Bayes Classifier* didalamnya. yang mana Pengujian performa terdiri dari perhitungan nilai akurasi, presisi, *recall*, dan *f-measure*. hasil pengujian ditunjukkan dalam bentuk *confusion matrix*.

Tabel 15. *Confusion matrix* hasil pengujian metode NBC

	Actual Terdiagnosa Penyakit Jantung	Actual Tidak Terdiagnosa Penyakit Jantung
Classified Terdiagnosa Penyakit Jantung	TP=3187	FN=333
Classified Tidak Terdiagnosa Penyakit Jantung	FP=2579	TN=901

Setelah diperoleh TP, FP, FN dan TN maka nilai akurasi, presisi, *recall*, *f-measure* dapat dihitung seperti yang ditunjukkan berikut:

$$AKURASI = \frac{(3187 + 901)}{(3187 + 2579 + 901 + 333)} = 0,58 = 58\%$$

$$PRESISI = \frac{3187}{3187 + 2579} = 0,55 = 55\%$$

$$RECALL = \frac{3187}{3187 + 333} = 0,90 = 90\%$$

$$F - measure = 2 \frac{0,58 \times 0,90}{0,58 + 0,90} = 0,68 = 68\%$$

3) Hasil pengujian performa

Berdasarkan hasil pengujian performa pada metode *K-Nearest Naighbor* dan *Naive Bayes Classifier* pada pengujian performa akurasi, presisi, *recall*, dan *f-measure*. Didapat perbandingan Performa metode klasifikasi terbaik pada dataset Penyakit jantung yaitu metode KNN dengan menggunakan pengukuran jarak manhattan diperoleh nilai performa terbaik pada data testing 20% yaitu akurasi sebesar 69%, presisi 66%, *recall* 75%, *f-measure* 70% pada nilai K=74.

IV. Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa Penerapan metode KNN pada dataset Penyakit jantung diperoleh nilai performa terbaik pada data *testing* 20% yaitu akurasi sebesar 68%, presisi 66%, *recall* 74%, *f-measure* 70% pada nilai K=236 dan metode jarak Manhattan. Penerapan metode KNN pada dataset Penyakit jantung diperoleh nilai performa terbaik pada data testing 20% yaitu akurasi sebesar 66%, presisi 65%, *recall* 70%, *f-measure* 67% pada nilai K=236 dan metode jarak *Euclidean*. Penerapan metode NBC pada dataset Penyakit jantung diperoleh nilai performa terbaik pada data *testing* 10% yaitu akurasi sebesar 58%, presisi 55%, *recall* 90%, *f-measure* 68%. Berdasarkan hasil performa perbandingan metode KNN dan NBC menunjukkan metode KNN memperoleh performa lebih baik dibandingkan dengan metode NBC.

Daftar Pustaka

- [1] A. Hadi, "Faktor Risiko Terjadinya Penyakit Jantung Koroner Pada Pasien Rumah Sakit Umum Meuraxa Banda Aceh (Risk factors of coronary heart disease in Meuraxa hospital of Banda Aceh)," *J. Action Aceh Nutr. Journal*, 2(1), 32–42, 2017.
- [2] D. Ramli and Y. Karan, "Anatomi dan Fisiologi Kompleks Mitral," *J. Kesehat. Andalas*, pp. 103–112, 2018.
- [3] A. A. Karim, H. Azis, and Y. Salim, "Salim, "Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana 1," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf*, vol. 3, pp. 84–87, 2018.
- [4] H. Azis, R. D. Mallongi, D. Lantara, and Y. Salim, "Comparison of Floyd-Warshall Algorithm and Greedy Algorithm in Determining the Shortest Route," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIconCIT*, pp. 294–298, 2018.
- [5] N. Fadhillah, H. Azis, and D. Lantara, "Validasi Pencarian Kata Kunci Menggunakan Algoritma Levenshtein Distance Berdasarkan Metode Approximate String Matching," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf*, vol. 1, pp. 3–7.
- [6] A. Fitria and H. Azis, "Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf*, vol. 3, pp. 102–106, 2018.
- [7] 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.
- [8] A. A. Karim, H. Azis, and Y. Salim, "Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana 1," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf*, vol. 3, no. 2, pp. 84–87, 2018.
- [9] 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.
- [10] Hasran, "Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 1–4, 2020.
- [11] D. Susanti, "Analisis Modifikasi Metode Playfiar Cipher Dalam Pengamanan Data," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 1–80, 2020.
- [12] H. Nursan and Muslim, "Penerapan Metode Digital Watermarking dan Privilege pada Dokumen Skripsi," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 19–22, 2020.
- [13] A. Maulida, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020.
- [14] F. T. Admojo and Ahsanawati, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020.