Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021
Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021
ISSN: 2598-0256, E-ISSN: 2598-0238

Komparasi Algoritma *Random Forest, Naïve Bayes* dan *K-Nearest Neighbor* Dalam klasifikasi Data Penyakit Jantung

Amril Samosir[1,*], Ms Hasibuan[2], Wahyu Eko Justino[3], Tri Hariyono[4]

[1,2,3,4]Institut Informatika dan Bisnis Darmajaya
Email: [1]*Amril5amos@gmail.com*, [2]*msaid@darmajaya.ac.id*, [3]*wahyu.justino.2021211030@mail.darmajaya.ac.id*, [4]*tri.haryono.2021211029@mail.darmajaya.ac.id*

Abstract

Heart disease is the leading cause of death after stroke and also can occur at any age. Computer-based methods encourage researchers to be one way for research on coronary heart disease. Processing large amounts of data can be done by classification using certain algorithms so that the results are fast and accurate. Commonly used classification methods include Naïve Bayes, K-Nearest Neighbor, and Random Forest. The Naïve Bayes method uses probability in each data, the K-Nearest Neighbor method uses distance calculations, while the Random Forest method uses several unified decision trees. This paper aims to compare the three algorithms in classifying heart disease data. Comparison of algorithms will be seen based on performance measures consisting of the level of accuracy, recall in each class, and precision in each class. Each algorithm is tested using cross validation. Based on the results of a comparison of 304 heart disease datasets, the Naïve Bayes algorithm is better and optimal than the K-Nearest Neighbor and Random Forest algorithms for classifying heart disease. The results of the classification using the Naïve Bayes algorithm have an average accuracy of 0.91 AUC, 0.84 CA, 0.84 F1, 0.839 Precision and 0.84 Recall.

Keywords: *Classification; machine learning; Naive Bayes; Random Forest, K-Nearest Neighbor; Dataset.*

Abstrak

Penyakit jantung merupakan penyabab kematian tertinggi setelah stroke dan dapat terjadi pada semua usia. Metode berbasis komputer mendorong peneliti untuk menjadi salah satu cara bagi penelitian terhadap penyakit jantung koroner. Pengolahan data dalam jumlah besar dapat dilakukan dengan klasifikasi menggunakan algoritma tertentu sehingga hasilnya cepat dan akurat. Metode klasifikasi yang umum digunakan antara lain *Naïve Bayes, K-Nearest Neighbor*, dan *Random Forest*. Metode *Naïve Bayes* menggunakan probabilitas disetiap data, metode *K-Nearest Neighbor* menggunakan perhitungan jarak, sedangkan metode *Random Forest* menggunakan beberapa pohon keputusan yang disatukan. Penulisan ini bertujuan untuk membandingkan ketiga algoritma tersebut dalam mengklasifikasikan data penyakit jantung. Perbandingan algoritma akan dilihat berdasarkan *performance measure* yang terdiri dari tingkatan akurasi, *recall* disetiap kelas, dan presisi disetiap kelas. Pada setiap algoritma diuji menggunakan *cross validation*. Berdasarkan hasil perbandingan terhadap 304 dataset penyakit jantung, algoritma *Naïve Bayes* lebih baik dan optimal dibanding dengan Algoritma, *K-Nearest Neighbor* dan *Random Forest* untuk mengklasifikasikan penyakit jantung. Hasil klasifikasi dengan algoritma *Naïve Bayes* memiliki rerata hasil akurasi sebesar 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 *Precision* dan 0,84 *Recall*.

Kata Kunci: *Klasifikasi;machine learning, Naïve Bayes, Random Forest, K-Nearest Neighbor, Dataset.*

1. PENDAHULUAN

Kesehatan merupakan faktor terpenting yang harus dijaga oleh setiap individu, agar dalam menjalani aktivitas sehari– hari menjadi lebih produktif, tidak mudah lelah dan lebih fokus dalam menyelesaikan suatu rutinitas dalam keseharian bagi setiap individu.

Sistem peredaran darah manusia adalah bagian sistem yang penting yang terjadi ditubuh manusia. Sistem peredaran darah ini memiliki dua fungsi utama, yakni untuk memberikan oksigen dan nutrisi keseluruh organ tubuh manusia serta mengangkut sisa dari hasil metabolisme. Dalam sistem peredaran darah manusia, organ tubuh yang penting adalah jantung. Jantung memiliki peranan sebagai alat pemompa darah yang mengalirkan darah keseluruh tubuh manusia. Jika organ jantung mengalami suatu gangguan atau kerusakan maka akan mengakibatkan terganggunya seluruh kinerja organ didalam tubuh manusia. Di Indonesia pada tahun 2014, hasil survei *Sample Regristration System* (SRS) menunjukkan bahwa penyakit jantung merupakan penyebab kematian tertinggi pada semua umur setelah stroke, yakni sebesar 12,9% [1].

ISSN: 2598-0256, E-ISSN: 2598-0238

Sebagian pasien yang menderita penyakit jantung mereka tidak mengetahui tanda-tanda gejala awal yang dirasakan dan tidak sedikit banyak pasien penderita penyakit jantung konorer yang meninggal dikarenakan serangan jantung. Kurangnya kesadaran pasien terhadap pola hidup sehat dan minimnya informasi penyakit jantung konorer yang dapat membuat seseorang tidak mengenali gejala awalnya dari penyakit ini. Cara untuk mendeteksi penyakit jantung dapat dilakukan dengan cara manual, yaitu dengan konsultasi langsung ke dokter spesialis jantung dan melakukan beberapa pemeriksaan laboratorium yang kemudian harus dikonsultasikan kembali oleh dokter spesialis jantung. Hal ini tentu saja memerlukan biaya yang relatif besar. Dengan adanya resiko kematian yang sangat tinggi, maka diperlukan suatu sistem yang dapat mendeteksi penyakit jantung koroner pada penderita secara akurat tetapi dengan biaya yang tidak besar.

Kasus ini mengundang banyak perhatian para peneliti, untuk dapat melakukan sebuah sistem yang akurat dengan harapan setelah adanya sistem ini dapat membantu pasien dan tidak memerlukan biaya yang cukup besar. Penelitian ini menggunakan sistem yang salah satunya menggunakan metode berbasis komputer. Metode ini banyak dikembangkan dengan bantuan komputasi cerdas yang mampu mengolah data dalam jumlah yang besar. Pengolahan data dalam jumlah besar dapat dilakukan dengan klasifikasi menggunakan algoritma tertentu sehingga hasilnya cepat dan akurat.

Metode klasifikasi yang biasanya digunakan antara lain *Naïve Bayes* [2], *K-Nearest Neighbor* [2, 3], *Random Forest* [4]. Pada penelitian ini, beberapa metode klasifikasi diimplementasikan pada kasus pengenalan penyakit jantung koroner untuk kemudian dapat dibandingkan hasil dari *performance measure* (akurasi, *recall*, dan presisi).

Beberapa penelitian yang membahas klasifikasi penyakit jantung koroner diantaranya penelitian yang dilakukan oleh Retnasari dan Rahmawati. Penelitian tersebut membandingkan dua algoritma, yakni algoritma *Naïve Bayes* dan algoritma *C4.5*. Pada hasil penelitian tersebut menunjukkan bahwa algoritma *Naïve Bayes* mendapatkan nilai akurasi 86,67%. Perbandingan Algoritma Klasifikasi algoritma *C4.5* mendapatkan nilai 83,70% [5].

Algoritma *Naïve Bayes*, *Random Forest*, *Decision Tree* dan *K-Nearest Neighbor* adalah algoritma pengklasifikasian dokumen teks yang umumnya mendapat nilai akurasi relatif tinggi. Hal ini dibuktikan melalui penelitian yang dilakukan oleh Dewi. Penelitian tersebut membandingkan lima algoritma, yakni algoritma *Neural Network*, *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor* dan *Logistic Regresion*. Hasil penelitian tersebut menunjukkan bahwa algoritma *Neural Network* mendapatkan nilai akurasi tertinggi yaitu 89,71%. Algoritma dengan nilai akurasi kedua tertinggi yaitu algoritma *Logistic Regression* yang mendapatkan nilai akurasi 89,32%, lalu algoritma *Decision Tree* mendapatkan nilai akurasi 89,10%. Selanjutnya, algoritma *K-Nearest Neighbor* dengan nilai akurasi 87,79% dan yang terakhir algoritma *Naïve Bayes* dengan nilai akurasi 84,70% [6].

Algoritma *Random Forest*, *Naivie Bayes* dan *K-Nearest Neighbor* masing-masing memiliki kelebihan dan kekurangan. Dengan demikian, pada penelitian ini dilakukan perbandingan algoritma diatas untuk memperoleh algoritma yang paling cocok dalam klasifikasi terhadap data penyakit jantung. Parameter pembanding ketiga algoritma tersebut yakni hasil *performance measure* (akurasi, *recall*, dan presisi).

2. KERANGKA TEORI

2.1 Naïve Bayes

Naïve Bayes merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naïve Bayes didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network . Teorema Bayes memiliki bentuk umum seperti:

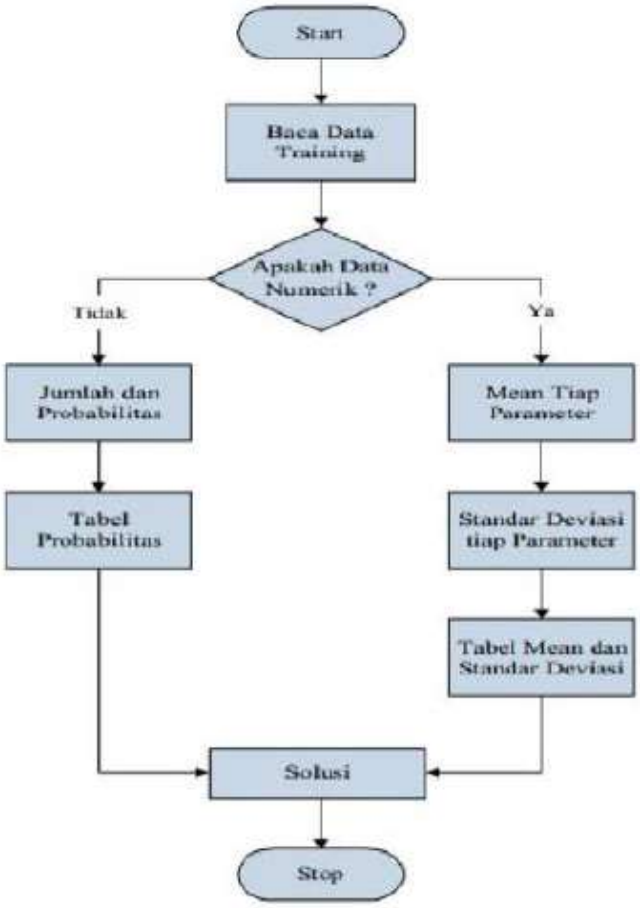
$$P(X|H) = \frac{P(H|X)P(X)}{P(H)}$$

Naïve Bayes merupakan penyederhanaan dari Teorema Bayes. Berikut ini merupakan rumus Naive Bayes setelah disederhanakan:

$$P(X|H) = P(H|X)P(X)$$

Keterangan:
W : data dengan *class* yang belum diketahui
H : hipotesis data X, merupakan suatu *class* spesifik
P(H|X) : probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
P(H) : probabilitas hipotesis H (*prior probability*)
P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H
P(X) : probabilitas dari X.

Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021
Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021
ISSN: 2598-0256, E-ISSN: 2598-0238



Gambar 1. Alur pemodelan Naive Bayes

2.2. *Algoritma Random Forest*

Pada proses perancangan model algoritma *Random Forest*, model algoritma *Random Forest* yang dibangun dengan menggunakan 'n' pohon *Decision Tree* dengan mencari nilai 'n' terbaik, maka didapatkan hasil optimum dari algoritma *Random Forest*, algoritma CART sebagai *criterion*nya, serta menggunakan *splitter best* untuk mendapatkan hasil akurasi terbaik dalam kasus data penyakit jantung .

2.3. Algoritma K-Nearest Neighbor

Proses perancangan model algoritma *K-Nearest Neighbor*, model algoritma *K-Nearest Neighbor* dibangun dengan menggunakan beberapa nilai k yang umum digunakan, kemudian dipilih nilai k yang menghasilkan performa terbaik. Ukuran jarak yang digunakan adalah *Euclidean Distance* dan *Manhattan Distance*, kemudian dibandingkan dan diambil serta dilihat model yang akan memberikan performa terbaik dalam kasus data penyakit jantung .

2*.4. Dataset*

Proses akuisisi data dilakukan dengan mengunduh dataset penyakit jantung koroner yang berasal dari lokasi *Cleveland Clinic Foundation* yang telah diadopsi oleh instansi *Hungarian Institute of Cardiology* di Budapest [7]. Proses pengolahan data untuk menghasillkan informasi menggunakan dataset jantung yang memilik 14 atribut seperti terlihat pada gambar berikut ini

ISSN: 2598-0256, E-ISSN: 2598-0238

	Name	Type	Role	Values
1	age	N numeric	feature	
2	sex	C categorical	feature	0, 1
3	cp	N numeric	feature	
4	trtbps	N numeric	feature	
5	chol	N numeric	feature	
6	fbs	C categorical	feature	0, 1
7	restecg	N numeric	feature	
8	thalachh	N numeric	feature	
9	exng	C categorical	feature	0, 1
10	oldpeak	N numeric	feature	
11	slp	N numeric	feature	
12	caa	N numeric	feature	
13	thall	N numeric	feature	
14	**output**	**C categorical**	**target**	**0, 1**

Gambar 2. Atribut Dataset jantung

Dari atribut yang dimiliki dataset jantung dapat diidentifikasi sebagai berikut: terdiri dari 13 atribut yang bersifat independen, 1 atribut yang bersifat dependen. 13 Atribut yang bersifat independen adalah sebagai berikut :

1. *Age*: Umur pasien.
2. *Sex*: Jenis kelamin pasien, atribut ini memiliki 2 nilai, yakni nilai 1 untuk laki-laki dan nilai 0 untuk perempuan.
3. Cp: Tipe nyeri dada yang diderita pasien. Atribut ini memilik 4 nilai, yaitu :
   Nilai 0: *asymptomatic*
   Nilai 1: *atypical angina*
   Nilai 2: *non anginal pain*
   Nilai 3: *typical angina*
4. Trestbps: *Resting blood pressure* yaitu tekanan darah pasien ketika dalam keadaan istirahat. Satuan yang dipakai adalah mm Hg.
5. Chol: *Cholesterol* yaitu kadar kolesterol dalam darah pasien, dengan satuan mg/dl.
6. Fbs: *Fasting blood sugar* yaitu kadar gula darah pasien, atribut fbs ini hanya memiliki 2 nilai yaitu 1 jika kadar gula darah pasien lebih dari 120 mg/dl, dan 0 jika kadar gula darah pasien kurang dari sama dengan 120 mg/dl.
7. Restecg: *Resting electrocardiographic* yaitu kondisi ECG pasien ketika dalam keadaan istirahat. Atribut ini memiliki 3 nilai yaitu nilai 1 untuk keadaan normal, nilai 2 untuk keadaan ST-T *wave abnormality* yaitu keadaan dimana gelombang *inversions* T dan atau ST meningkat maupun menurun lebih dari 0,5 mV dan nilai 3 untuk keadaan dimana ventricular kiri mengalami hipertropi.
8. Thalach: Rata-rata detak jantung pasien dalam satu menit.
9. Exang: Keadaan dimana pasien akan mengalami nyeri dada apabila berolah raga, 0 jika tidak nyeri, dan 1 jika menyebabkan nyeri.
10. Oldpeak: Penurunan ST akibat olahraga.
11. Slope: Slope dari puncak ST setelah berolah raga. Atribut ini memiliki 3 nilai yaitu 0 untuk *downsloping*, 1 untuk flat, dan 2 untuk *upsloping*.
12. Ca: Banyaknya pembuluh darah yang terdeteksi melalui proses pewarnaan *flourosopy*.
13. Thal: Detak jantung pasien. Atribut ini memiliki 3 nilai yaitu 1 untuk fixed defect, 2 untuk normal dan 3 untuk reversal defect.

ISSN: 2598-0256, E-ISSN: 2598-0238

Variabel dependen pada penelitian ini yaitu variabel target yang dimaksudkan yaitu *output*, yang berarti hasil diagnosa penyakit jantung, yang memiliki 2 nilai, yakni 0 untuk terdiagnosa positif terkena penyakit jantung koroner, dan 1 untuk negatif terkena penyakit jantung.

2.5. Implementasi Dataset

Penerapan data ini dilakukan dengan pemisahan dataset menjadi dua bagian, yakni data *training* dan data *testing*. Jumlah data *training* dan data *testing* yang digunakan adalah perbandingan. Pada penelitian ini terdapat 304 data pasien jantung, sehingga jumlah data *training* dan data *testing* berturut-turut adalah 294 data dan 10 data. Data *training* digunakan untuk melatih *classifier* dalam mengenali karakteristik pasien yang positif terkena jantung maupun yang negatif. Data *testing* dapat digunakan dalam uji coba, terhadap model klasifikasi yang dihasilkan dan menentukan performa dari model klasifikasi dengan cara membandingkan hasil klasifikasi model tiap data dalam data *testing* dengan label sebenarnya.

2.6. Uji Coba Model

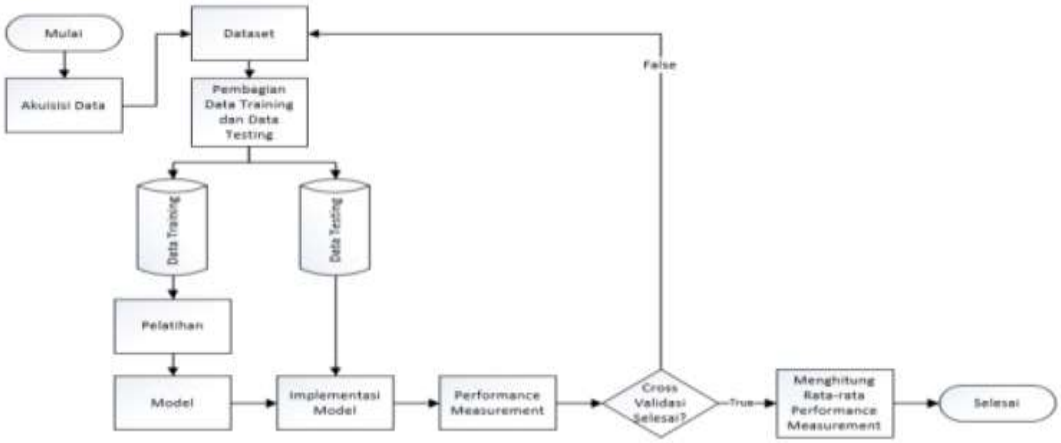
Pengujian dilakukan terhadap model-model yang dihasilkan dari 3 buah algoritma, yaitu algoritma *Naïve Bayes*, algoritma *K-Nearest Neighbor*, dan algoritma *Random Forest*, yang menghasilkan *performance measure* yang berbeda-beda terhadap kasus data penyakit jantung . Berdasarkan prinsip *K-fold cross validation*, maka *performance measure* yang diberikan adalah nilai rata-rata performa dari tiap model yang dihasilkan dari k percobaan terhadap masing-masing model tersebut. Adapun *performance measure* yang digunakan adalah *recall*, presisi, dan akurasi yang disajikan dalam bentuk persentase dengan rumusan masing-masing berdasarkan nilai-nilai pada *confusion matrix*.

3. METODOLOGI

Pada tahapan ini dilakukan pendekatan dengan komparasi 3 algoritma yaitu *Naïve Bayes, Random Forest* dan *K-NN* dengan menggunakan perangkat lunak yaitu *Orange* dan menggunakan dataset penderita sakit jantung sebagi data yang akan menghasilkan beberapa informasi atau *output* dari yang dinginkan.

3.1 Kerangka Kerja

Pada tahapan ini akan diberikan gambaran dari kerangka kerja dalam bentuk flowchart yang akan memberikan informasi dari tahapan-tahapan yang akan digunakan untuk implementasi dan menganalisa parameter yang akan digunakan untuk memproses data dengan melakukan pembagian dataset yang telah diolah. Pembagian data dilakukan dengan membagi data utama menjadi dua bagian, yaitu data training dan data testing, untuk data testing sebanyak 294 record sedangkan untuk data testing 10 record, hal ini dilakukan agar bisa memprediksi dari ketiga algoritma yang digunakan yaitu algoritma *Naïve Bayes, K-Nearest Neighbor dan Random Forest*. Gambaran umum dari tahapan tersebut dapat dilihat pada gambar 1.



Gambar 1. Skema Umum Penelitian

ISSN: 2598-0256, E-ISSN: 2598-0238

3.2 Analisis Data

Proses ini bertujuan untuk menyusun data secara sistematis yang akan menghasilkan informasi-informasi yang terbentuk menjadi dataset dengan atribut-atribut yang saling mendukung antara satu dengan yang lain nya. Adapun dataset yang digunakan dalam penulisan ini seperti terlihat pada tabel 1 berikut :

Tabel 1. Dataset Jantung

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
…													
56	1	1	130	221	0	0	163	0	0	2	0	3	1
56	1	1	120	240	0	1	169	0	0	0	0	2	1
45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
68	1	0	144	193	1	1	141	0	3.4	1	2	3	0

Dalam penulisan ini, penulis menggunakan komparasi Algoritma yaitu Algoritma Naive Bayes, K-NN dan Random forest, untuk melakukan prediksi terhadap ketiga Algoritma tersebut dibutuhkan data training dan data testing seperti terlihat pada Tabel 2 dan tabel 3

Tabel 2. Data Training 294 Record

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
…													
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1

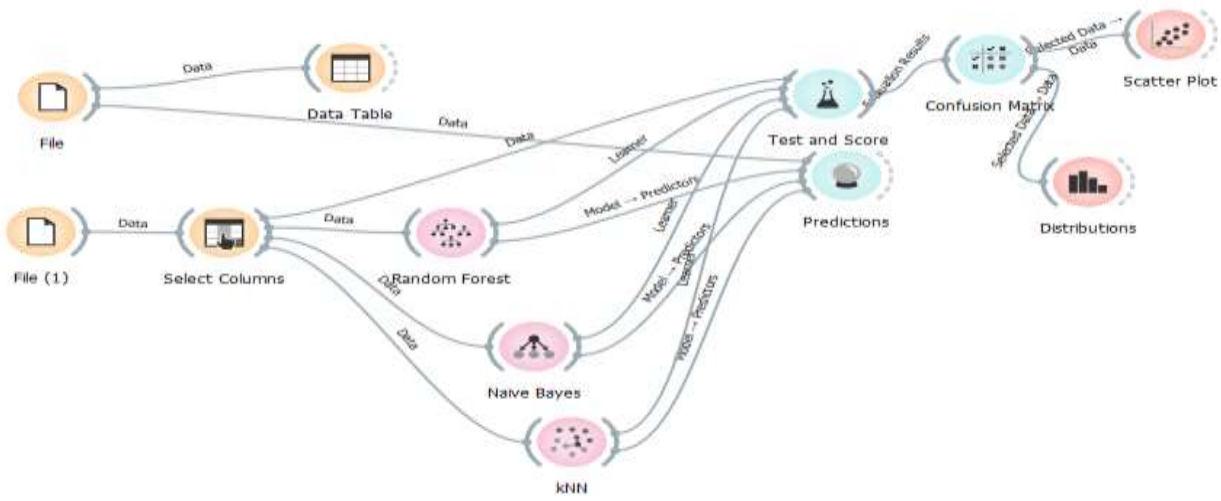
Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021
Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021
ISSN: 2598-0256, E-ISSN: 2598-0238

Tabel 3. Data Testing 10 Record

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
67	1	2	152	212	0	0	150	0	0.8	1	0	3	0
44	1	0	120	169	0	1	144	1	2.8	0	0	1	0
63	1	0	140	187	0	0	144	1	4	2	2	3	0
63	0	0	124	197	0	1	136	1	0	1	0	2	0
59	1	0	164	176	1	0	90	0	1	1	2	1	0
57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
56	1	1	130	221	0	0	163	0	0	2	0	3	1
56	1	1	120	240	0	1	169	0	0	0	0	2	1
45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
68	1	0	144	193	1	1	141	0	3.4	1	2	3	0

3.3 Proses Analis

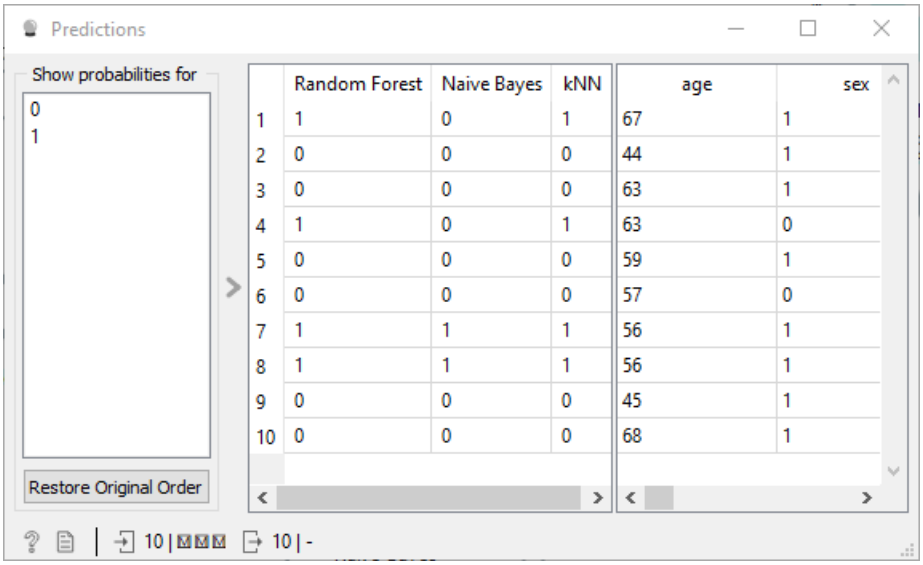
Tahapan ini penulis akan mendesain Algoritma untuk memprediksi ke tiga Algoritma tersebut kedalam tools orange sebagai wadah untuk menjalankan komparasi Algoritma Naive Bayes, K-NN dan Random Forest, desain dari komponen-komponen seperti terlihat dalam Gambar 1 workflow komparasi Algoritma sebagai berikut :



Gambar 1. Komparasi Algoritma

Proses komparasi Algoritma Naive Bayes, K-NN dan Random Forest menghasilkan nilai prediksi seperti terlihat pada gambar 2 berikut ini :

ISSN: 2598-0256, E-ISSN: 2598-0238



Gambar 2 Prediksi Ketiga Algoritma

3.4 Hasil Analis

Proses hasil akhir dari komparasi Algoritma yaitu Naivie Bayes, K-NN dan Random Forest memiliki hasil yang berbeda-beda, hal ini disebakan masing-masing algoritma memiliki kelebihan dan kekurangan dalam memproses data atau hasil akhir, seperti dapat dilihat dari hasil tes skor ketiga algoritma dengan menggunakan Testscore seperti terlihat pada tabel 4 berikut ini:

Tabel 4 Akurasi dari kompari Algoritma

Model	*AUC*	*CA*	*F1*	*Precision*	*Recall*
k-NN	0,686	0,645	0,641	0,642	0,645
Random Forest	0,899	0.809	0,808	0,809	0,809
Naïve Bayes	0,910	0,840	0,840	0,839	0,840

4. HASIL DAN PEMBAHASAN

Proses yang dihasilkan dari pengujian komparasi tiga Algoritma memberikan hasil terhadap: prediksi, akurasi dan *confusion matrix*. Hasil dari prediksi ketiga algoritma yang memiliki nilai paling benar yaitu pada *Algoritma Naïve Bayes*, dengan nilai *0,0,0,0,0,0,1,1,0,0*, artinya sangat sesuai dengan nilai data testing yaitu *: 0,0,0,0,0,0,1,1,0,0* utntuk *Algoritma K-NN* dengan nilai *1,0,0,1,0,0,1,1,0,0* memiliki *2 nilai kesalahan, yaitu nilai 1* pada baris pertama dan nilai 1 pada baris ke empat dan *Random Forest dengan nilai 1,0,0,1,0,0,1,1,0,0* memiliki *2* nilai kesalahan, yaitu nilai *1* pada baris pertama dan nilai *1* pada baris ke empat. Pengujian sample dataset sebanyak 304 record di uji menggunakan tiga Algoritma, dari pengujian ini diperoleh hasil akurasi yang terbaik dari ke tiga algoritma yang digunakan. Adapun perolehan hasil komparasi dari ke tiga Algoritma menghasilkan nilai skor sebesar : *Algoritma Naive Bayes sebesar 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 Precision dan 0,84 Recall*, hasil akurasi yang diperoleh *algoritma Random Forest* sebesar *0,884 AUC, 0,788 CA, 0,787 F1, 0,788 Precision , 0,788 Recall* dan *Algoritma kNN memiliki akurasi* sebesar *0,686 AUC, 0,645 CA, 0,641 F1, 0,642 Precision, 0,645 Recall.*

Hasil tersebut menunjukkan bahwa model klasifikasi algoritma *Naïve Bayes* menghasilkan *performance measure* yang terbaik untuk mengklasifikasikan penyakit jantung dibandingkan dengan model klasifikasi algoritma *Random Forest* dan Algoritma *K-Nearest Neighbor*.

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa implementasi dan perbandingan algoritma *Naïve Bayes, K-Nearest Neighbor* dan *Random Forest* terhadap data kasus penyakit jantung berhasil

ISSN: 2598-0256, E-ISSN: 2598-0238

direalisasikan. Implementasi dan uji coba telah dilakukan 304 dataset penyakit jantung. Selain itu, telah dilakukan perbandingan terhadap hasil uji coba Algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Random Forest*.
Berdasarkan perbandingan hasil uji coba, *performance measure* algoritma *Naïve Bayes* memiliki hasil yang lebih baik dibanding dengan algoritma, *K-Nearest Neighbor* dan *Random Forest* dengan metode *K-Fold Cross Validation*. Algoritma *Naïve Bayes* dapat memberikan rerata hasil akurasi sebesar 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 *Precision* dan 0,84 *Recall*.

Dengan demikian dapat dikatakan bahwa algoritma *Naïve Bayes* adalah algoritma yang terbaik dalam mengklasifikasi kasus penyakit jantung dibanding dengan algoritma, *K-Nearest Neighbor* dan *Random Forest.*
Pada penelitian ini dilakukan uji coba menggunakan data dalam jumlah yang kecil, sehingga untuk penelitian selanjutnya dapat mengembangkan dengan melakukan uji coba pada data dengan jumlah yang lebih besar. Pada penelitian lebih lanjut juga dapat menambahkan algoritma klasifikasi lainnya sehingga mendapat hasil perbandingan yang lebih beragam. Selain itu, penelitian lebih lanjut dapat mengembangkan klasifikasi kelas yang digunakan, agar menjadi rinci dalam mengklasifikasi penyakit jantung koroner berdasarkan level atau tingkatan tertentu.

UCAPAN TERIMA KASIH

Teruntuk temen-temen penulis yang sudah banyak berkontribusi dari sisi waktu dan pikiran, sehingga lahirlah naskah ini

DAFTAR PUSTAKA

Kementerian Kesehatan RI, “Penyakit jantung penyebab kematian tertinggi, kemenkes ingatkan cerdik,” *Kementerian Kesehatan RI*, 2017. [Daring]. Tersedia: http://www.depkes.go.id/article/view/17073100005/penyakit-jantung-penyebab-kematian-tertinggi-kemenkes-ingatkan-cerdik-.html. [Diakses: 25 April 2019].

D. Sartika dan D. I. Sensuse, “Perbandingan algoritma klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada studi kasus pengambilan keputusan pemilihan pola pakaian,” Jatisi, vol. 1, no. 2, hal. 153 – 154, 2017.

M. Lestari, “Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung,” *Faktor Exacta*, vol. 7, no. 4, hal. 366 – 371, 2014.

M. R. Amiarrahman dan T. Handhika, “Analisis dan implementasi algoritma klasifikasi Random Forest dalam pengenalan Bahasa Isyarat Indonesia (BISINDO),” Prosiding Seminar Nasional Inovasi Teknologi, 2017, hal. 83 – 88.

T. Retnasari dan E. Rahmawati, “Diagnosa prediksi penyakit jantung dengan model algoritma Naïve Bayes dan algoritma C4.5,” Prosiding Konferensi Nasional Ilmu Sosial dan Teknologi (KNiST), 2017, hal. 7 – 12.

S. Dewi, “Komparasi 5 metode algoritma klasifikasi data mining pada prediksi keberhasilan pemasaran produk layanan perbankan,” *Jurnal Techno Nusa Mandiri*, vol. 8, no. 1, hal. 60 – 66, 2016.

Kaggle Dataset, “Heart disease UCI,” *Kaggle Dataset*, 2019. [Daring]. Tersedia: https://www.kaggle.com/ronitf/heart-disease-uci. [Diakses: 25 April 2021].