

# PERBANDINGAN ALGORITMA KLASIFIKASI DALAM PENGKLASIFIKASIAN DATA PENYAKIT JANTUNG KORONER

<sup>1</sup>Ardea Bagas Wibisono, <sup>2</sup>Achmad Fahrurozi

<sup>1,2</sup>Fakultas Teknologi Industri Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

<sup>1</sup>ardeabagas22@gmail.com, <sup>2</sup>achmad\_fahrurozi@staffsite.gunadarma.ac.id

## Abstrak

*Penyakit Jantung Koroner (PJK) menjadi penyebab kematian tertinggi pada semua umur setelah stroke. Hal ini mendorong banyak penelitian terhadap penyakit jantung koroner, salah satunya menggunakan metode berbasis komputer. 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, Decision Tree dan Random Forest. Metode Naïve Bayes menggunakan probabilitas disetiap data, metode K-Nearest Neighbor menggunakan perhitungan jarak, metode Decision Tree menggunakan pohon keputusan, sedangkan metode Random Forest menggunakan beberapa pohon keputusan yang disatukan. Penelitian ini bertujuan untuk membandingkan keempat algoritma tersebut dalam mengklasifikasikan data penyakit jantung koroner. 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 300 dataset penyakit jantung koroner, algoritma Random Forest lebih baik dan optimal dibanding dengan Algoritma Naïve Bayes, K-Nearest Neighbor, dan Decision Tree untuk mengklasifikasikan penyakit jantung koroner. Hasil klasifikasi dengan algoritma Random Forest memiliki rerata tingkat akurasi sebesar 85,668 % dengan recall kelas '1' adalah 89 %, recall kelas '0' adalah 83,6%, presisi kelas '1' adalah 85%, dan presisi kelas '0' adalah 85,8%.*

**Kata Kunci:** *Klasifikasi, machine learning, penyakit jantung koroner.*

## Abstract

*Coronary Artery Disease (CAD) is the highest cause of death in all age categories after stroke. This fact encourages numerous studies on coronary artery disease, one of which using computer-based methods. Processing big data can be done by using classification with certain algorithms to obtain fast and accurate results. The commonly used classification methods include Naïve Bayes, K-Nearest Neighbor, Decision Tree and Random Forest. The Naïve Bayes method uses probability in each data, the K-Nearest Neighbor uses distance calculation, the Decision Tree uses a decision tree, and the Random Forest uses several integrated decision trees. This study aims to compare the four algorithms in classifying coronary artery disease data. Algorithm comparison was determined based on performance measure of each algorithm, which consists of accuracy level, recall in each class, and precision in each class. Each algorithm was tested using cross validation. Based on the comparison results of 300 coronary artery disease data, the Random Forest algorithm was better and more optimum compared to the Naïve Bayes Algorithm, K-Nearest Neighbor, and Decision Tree in classifying coronary artery disease. Classification results with the Random Forest algorithm have an average accuracy rate of 85.668% based on the following data: the class '1' recall was 89%, the class '0' recall was 83.6%, the class '1' precision was 85%, and the class '0' precision was 85.8%.*

**Keywords:** *classification, coronary artery disease, machine learning.*

## PENDAHULUAN

Sistem peredaran darah manusia merupakan salah satu sistem yang penting pada tubuh manusia. Sistem ini memiliki dua fungsi utama, yakni untuk mengedarkan oksigen dan nutrisi keseluruh organ tubuh manusia serta mengangkut sisa hasil metabolisme. Salah satu organ yang penting dalam sistem peredaran darah manusia adalah jantung. Jantung bertugas sebagai alat pompa untuk mengedarkan darah keseluruh tubuh manusia. Jika jantung mengalami suatu gangguan atau kerusakan maka akan mengakibatkan terganggunya seluruh kinerja organ didalam tubuh manusia. Menurut survei *Sample Regristration System (SRS)* pada 2014 di Indonesia menunjukkan bahwa Penyakit Jantung Koroner (PJK) menjadi penyebab kematian tertinggi pada semua umur setelah stroke, yakni sebesar 12,9% [1].

Kebanyakan pasien penderita penyakit jantung tidak mengetahui gejala-gejala awal yang dirasakan dan tidak sedikit banyak penderita penyakit jantung koroner yang meninggal disebabkan oleh serangan jantung. Masih kurangnya kesadaran terhadap pola hidup yang sehat dan kurangnya informasi penyakit jantung koroner yang dapat membuat seseorang tidak dapat mengenali gejala awalnya. Proses untuk mendeteksi penyakit jantung dapat dilakukan dengan cara manual, yakni 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 resiko kematian yang sangat tinggi, maka diperlukan suatu sistem yang dapat mendeteksi penyakit jantung koroner pada penderita secara akurat serta dengan biaya yang tidak besar.

Hal ini mendorong banyak penelitian terhadap penyakit jantung koroner, 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 umum digunakan antara lain *Naïve Bayes* [2], *K-Nearest Neighbor* [2, 3], *Decision Tree* [2, 4], *Random Forest* [5] dan *Support Vector Machines* [6]. Pada penelitian ini, beberapa metode klasifikasi diimplementasikan pada kasus pengenalan penyakit jantung koroner untuk kemudian dibandingkan hasil *performance measure* (akurasi, *recall*, dan presisi).

Beberapa penelitian yang membahas klasifikasi penyakit jantung koroner diantaranya penelitian 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% dan

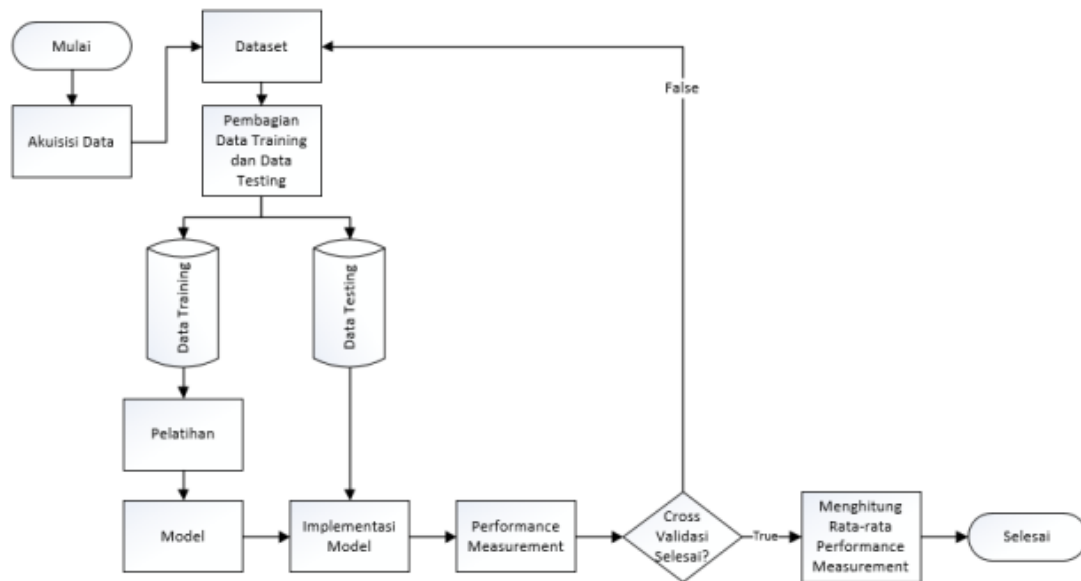
algoritma *C4.5* mendapatkan nilai 83,70% [7].

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 Regresion* 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% [8].

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

## METODE PENELITIAN

Pada metode penelitian terdiri atas beberapa tahapan proses, yakni mengakuisisi data yang akan digunakan untuk implementasi lalu menganalisa parameter yang digunakan untuk memproses data dengan melakukan pembagian data yang telah diolah. Pembagian data untuk data *training* dan data *testing* dengan perbandingan sebesar 80:20. Kemudian merancang sebuah model algoritma *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree* dan *Random Forest*. Setelah itu melakukan proses pelatihan data dengan algoritma yang telah dipilih. Pada tahap ini model akan mempelajari *feature* yang didapat dan melakukan pengubahan nilai parameter selama latihan berlangsung. Tahap berikutnya yakni mempelajari model dengan data uji yang kemudian data uji akan diklasifikasikan hasilnya dan dibandingkan dengan nilai parameter data uji yang sesungguhnya sehingga mendapatkan tingkat akurasi dalam klasifikasi proses uji data dan data uji yang sesungguhnya. Percobaan akan dilakukan dengan menggunakan *cross validation*, sehingga membutuhkan beberapa percobaan, dalam penelitian ini dilakukan 5 percobaan dalam satu algoritma. Hasil akhir dengan menggunakan *cross validation* dinyatakan dari rata-rata hasil akurasi dari 5 percobaan yang diterapkan pada setiap algoritma. Gambaran umum mengenai alur proses dari penelitian dapat dilihat pada Gambar 1.



Gambar 1. Skema Umum Penelitian

### Akuisisi Data

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 [9]. Data yang diambil terdiri dari 300 data pasien jantung koroner, di mana masing-masing data memiliki 14 variabel, terdiri atas 13 variabel independen dan 1 variabel dependen yang merupakan parameter hasil diagnosis penyakit jantung koroner (memiliki 2 buah nilai, yakni nilai 0 untuk positif dan nilai 1 untuk negatif).

### Dataset

Pada *dataset* penyakit jantung koroner terdapat 14 variabel yang terbagi menjadi 13 independen dan 1 variabel dependen. Variabel independen 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 memiliki 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*.

Variabel dependen pada penelitian ini yaitu variabel target yang berarti hasil diagnosa penyakit jantung, memiliki 2 nilai, yakni 0 untuk terdiagnosa positif terkena

penyakit jantung koroner, dan 1 untuk negatif terkena penyakit jantung koroner.

### **Pembagian Data Training dan Data Testing**

Pada tahap ini dilakukan pemisahan *dataset* menjadi dua bagian, yakni data *training* dan data *testing*. Jumlah data *training* dan data *testing* yang digunakan adalah perbandingan 80:20. Pada penelitian ini terdapat 300 data pasien jantung koroner, sehingga jumlah data *training* dan data *testing* berturut-turut adalah 240 data dan 60 data. Data *training* digunakan untuk melatih *classifier* dalam mengenali karakteristik pasien yang positif terkena jantung koroner maupun yang negatif. Data *testing* digunakan dalam uji coba terhadap model klasifikasi yang dihasilkan dan menentukan performa dari model klasifikasi dengan cara membandingkan hasil klasifikasi model terhadap tiap data dalam data *testing* dengan label sebenarnya.

### **Perancangan Model**

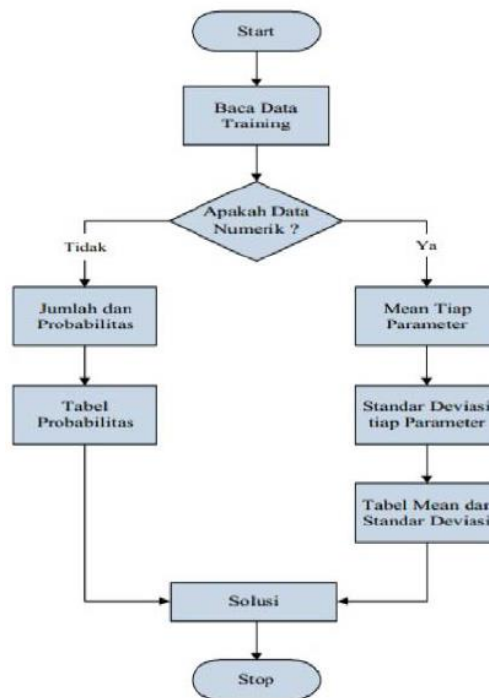
Perancangan model algoritma *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree* dan *Random Forest* yang dibangun akan digunakan sebagai *classifier* untuk mengelompokkan data yang positif terkena penyakit jantung koroner atau tidak berdasarkan parameter yang diberikan.

#### **1. Perancangan Model Algoritma Naïve Bayes**

Perancangan model algoritma *Naïve Bayes* tidak memiliki banyak pengaturan karena algoritma *Naïve Bayes* merupakan algoritma pengklasifikasian sederhana yang meng-

hitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Dalam penelitian ini digunakan distribusi frekuensi Normal dalam model perhitungan

probabilitasnya, sehingga data penyakit jantung koroner yang dihimpun diasumsikan berdistribusi Normal standar. Gambar 2 merupakan alur pembentukan model *Naïve Bayes*.



Gambar 2. Alur Model *Naïve Bayes*

## 2. Algoritma *k-Nearest Neighbor*

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

## 3. Algoritma *Decision Tree*

Pada proses perancangan model algoritma *Decision Tree*, model algoritma *Decision Tree* yang dibangun menggunakan algoritma *CART* sebagai *criterion*-nya, serta menggunakan *splitter best* untuk mendapatkan hasil akurasi terbaik dalam kasus data penyakit jantung koroner.

## 4. 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 *criterionnya*, serta menggunakan *splitter best* untuk mendapatkan hasil akurasi terbaik dalam kasus data penyakit jantung koroner.

### **Pelatihan Model**

Tahap selanjutnya adalah proses pelatihan model dengan menggunakan algoritma *Naïve Bayes*, *k-Nearest Neighbor*, *Decision Tree*, serta *Random Forest*. Proses pelatihan ini dilakukan dengan prinsip *k-fold cross validation*, yaitu dengan cara membagi data menjadi sejumlah *k* himpunan bagian, sedemikian sehingga setiap kali salah satu himpunan bagian digunakan sebagai data *testing*, maka  $k - 1$  himpunan bagian sisanya disatukan untuk membentuk data *training*. Estimasi kesalahan dirata-rata pada tiap uji coba, di mana uji coba dilakukan berulang sebanyak *k* kali, sedemikian sehingga masing-masing himpunan bagian dari *k* himpunan bagian tersebut berperan menjadi himpunan *testing* sebanyak tepat satu kali. Hal ini dilakukan untuk mendapatkan efektivitas total dari *dataset* yang ada. Pengulangan data latih dan data uji ini secara signifikan mengurangi varians karena sebagian besar data juga digunakan dalam data uji.

### **Pengujian Model**

Pengujian dilakukan terhadap model-model yang dihasilkan dari 4 buah algoritma, yaitu algoritma *Naïve Bayes*, algoritma *k-*

*Nearest Neighbor*, algoritma *Decision Tree* dan algoritma *Random Forest*, yang menghasilkan *performance measure* yang berbeda-beda terhadap kasus data penyakit jantung koroner. 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*.

## **HASIL DAN PEMBAHASAN**

Pengujian terhadap empat algoritma, yakni algoritma *Naïve Bayes*, algoritma *K-Nearest Neighbor*, algoritma *Decision Tree* dan algoritma *Random Forest* menghasilkan akurasi yang berbeda-beda terhadap kasus penyakit jantung koroner. Pengujian menggunakan 300 jumlah data dari *dataset* yang disediakan. Setiap data diuji terhadap seluruh data yang ada. Kemudian hasil pengujian tersebut diukur untuk mendapatkan tingkat ketepatan dalam mengklasifikasikan penyakit jantung koroner dengan menghitung rata-rata *recall*, presisi, dan akurasi dari setiap percobaan menggunakan *cross validation* yang disajikan dalam bentuk persentase. Tabel 1 menunjukkan rangkuman dari hasil *performance measure* terhadap seluruh hasil pengujian dalam masing-masing algoritma.

Tabel 1. Hasil *Performance Measure* Klasifikasi Penyakit Jantung Koroner

Algoritma	<i>Performance Measure</i>				
	<i>Recall '1'</i>	<i>Recall '0'</i>	<i>Presisi '1'</i>	<i>Presisi '0'</i>	<i>Akurasi</i>
<i>Naïve Bayes</i>	84,8 %	78,4 %	80,8 %	81,2 %	80,33 %
<i>K-Nearest Neighbor</i>	73,4 %	68 %	72,2 %	68,2 %	69,67 %
<i>Decision Tree</i>	82,8 %	78,2 %	81,2 %	79,4 %	80,33 %
<i>Random Forest</i>	89 %	83,6 %	85 %	85,8 %	85,668 %

Hasil *performance measure* pada Tabel 1 menunjukkan akurasi paling optimal untuk keempat algoritma. Berdasarkan hasil perbandingan pada Tabel 1 dapat disimpulkan bahwa hasil akurasi tertinggi terhadap klasifikasi penyakit jantung koroner diperoleh algoritma *Random Forest* sebesar 85,67 % dengan *recall* kelas '1' adalah 89 %, *recall* kelas '0' adalah 83,6%, presisi kelas '1' adalah 85%, dan presisi kelas '0' adalah 85,8%. Hasil akurasi yang diperoleh algoritma *Naïve Bayes* sebesar 80,33 % dengan *recall* kelas '1' adalah 84,8%, *recall* kelas '0' adalah 78,4%, presisi kelas '1' adalah 80,8%, dan presisi kelas '0' adalah 81,2%. Algoritma *Decision Tree* memiliki akurasi sebesar 80,33% dengan *recall* kelas '1' adalah 82,8%, *recall* kelas '0' adalah 78,2%, presisi kelas '1' adalah 81,2%, dan presisi kelas '0' adalah 79,4%. Hasil akurasi yang terendah yaitu algoritma *K-Nearest Neighbor* dengan akurasi sebesar 69,67% dengan *recall* kelas '1' adalah 73,4%, *recall* kelas '0' adalah 68%, presisi kelas '1' adalah 72,2%, dan presisi kelas '0' adalah 68,2%. Hasil tersebut menunjukkan bahwa model klasifikasi algoritma *Random Forest* menghasilkan *performance measure* yang terbaik untuk mengklasifikasikan penyakit

jantung koroner dibandingkan dengan model klasifikasi algoritma *Naïve Bayes*, Algoritma *Decision Tree* dan Algoritma *K-Nearest Neighbor*.

## KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa implementasi dan perbandingan algoritma *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree* dan *Random Forest* terhadap data kasus penyakit jantung koroner berhasil direalisasikan. Implementasi dan uji coba telah dilakukan 300 *dataset* penyakit jantung koroner. Selain itu, telah dilakukan perbandingan terhadap hasil uji coba Algoritma *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree* dan *Random Forest*.

Berdasarkan perbandingan hasil uji coba, *performance measure* algoritma *Random Forest* memiliki hasil yang lebih baik dibanding dengan algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Decision Tree* dengan metode *k-fold cross validation*. Algoritma *Random Forest* dapat memberikan rerata hasil akurasi sebesar 85,668 % dengan *recall* kelas '1' adalah 89 %, *recall* kelas '0' adalah 83,6%, presisi kelas '1' adalah 85%,



dan presisi kelas '0' adalah 85,8%. Dengan demikian dapat dikatakan bahwa algoritma *Random Forest* adalah algoritma yang terbaik dalam mengklasifikasi kasus penyakit jantung koroner dibanding dengan algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Decision Tree*.

Pada penelitian ini dilakukan uji coba menggunakan data dalam jumlah kecil, sehingga pada penelitian selanjutnya dapat dikembangkan dengan melakukan uji coba pada data dengan jumlah yang lebih besar. Pada penelitian lebih lanjut dapat juga menambahkan algoritma klasifikasi lainnya sehingga mendapat perbandingan yang lebih beragam. Selain itu, penelitian lebih lanjut juga dapat mengembangkan klasifikasi kelas yang digunakan, agar menjadi rinci dalam mengklasifikasi penyakit jantung koroner berdasarkan level atau tingkatan tertentu.

## DAFTAR PUSTAKA

- [1] Kementerian Kesehatan RI, "Penyakit jantung penyebab kematian tertinggi, kemenkes ingatkan cerdas," *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].
- [2] D. Sartika dan D. I. Sensuse, "Perbandingan algoritma klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada studi kasus pengambilan keputusan pemilihan pola pakaian," *Jatiji*, vol. 1, no. 2, hal. 153 – 154, 2017.
- [3] M. Lestari, "Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung," *Faktor Exacta*, vol. 7, no. 4, hal. 366 – 371, 2014.
- [4] F. Maspiyanti dan J. Gatc, "Diagnosa penyakit jantung pada ponsel menggunakan pohon keputusan," *Jurnal Teknologi Terpadu*, vol. 1, no. 1, hal. 13 – 20, 2015.
- [5] 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.
- [6] A. M. Puspitasari, D. E. Ratnawati, dan A. W. Widodo, "Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, hal. 802 – 810, 2018.
- [7] 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.

- [8] 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.
- [9] Kaggle Dataset, “Heart disease UCI,” *Kaggle Dataset*, 2019. [Daring]. Tersedia: <https://www.kaggle.com/ronitf/heart-disease-uci>. [Diakses: 22 April 2019].