

# Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung

Jalis Dwi Muthohhar<sup>1</sup>, Agus Prihanto<sup>2</sup>

<sup>1,2</sup> Jurusan Teknik Informatika/Teknik Informatika, Universitas Negeri Surabaya

<sup>1</sup>[jalis.17051204048@mhs.unesa.ac.id](mailto:jalis.17051204048@mhs.unesa.ac.id)

<sup>2</sup>[agusprihanto@unesa.ac.id](mailto:agusprihanto@unesa.ac.id)

**Abstrak**— Penyakit jantung menjadi penyebab tertinggi kematian didunia setelah *stroke*. Dengan kemajuan teknologi yang sangat pesat mendorong berbagai penelitian untuk mempermudah dalam mendiagnosa penyakit jantung menggunakan klasifikasi *machine learning*. Model klasifikasi yang digunakan yaitu *Decision Tree*, *Naive Bayes* dan *Random Forest Classifier*. Model *Decision Tree* menggunakan pohon keputusan, metode *Naive Bayes* menggunakan probabilitas disetiap data, sedangkan model *Random Forest Classifier* menggunakan sekumpulan pohon keputusan yang disatukan. Data yang digunakan adalah *Heart Attack Analysis & Prediction Dataset* diambil dari Kaggle. Penelitian ini bertujuan untuk mendapatkan model yang terbaik untuk mengklasifikasikan data penyakit jantung dari ketiga model tersebut. Untuk mendapatkan model yang terbaik penelitian ini melakukan *cross validation* pada setiap model *machine learning*. Dari hasil *cross validation* didapatkan adanya *overfitting* pada setiap model. Untuk mengatasi *overfitting* peneliti melakukan *fine tuning* pada setiap model dengan parameter *F1-Score*. Berdasarkan rata-rata hasil pengujian model *decision tree* menggunakan evaluasi *random search* mendapatkan nilai 0.844 dan *grid search* 0.84. Kemudian hasil pengujian model *naive bayes* tidak ada perbedaan antara evaluasi menggunakan *random search* maupun *grid search* yaitu sebesar 0.85. Selanjutnya hasil pengujian model *random forest classifier* menggunakan evaluasi *random search* mendapatkan nilai 0.852 dan *grid search* 0.868. Dengan demikian dapat dilihat dari hasil pengujian bahwa model *random forest classifier* merupakan model terbaik dalam mengklasifikasi penyakit jantung baik itu menggunakan *hyperparameter tuning grid search*.

**Kata Kunci**— *Machine Learning, Klasifikasi, Decision Tree, Naive Bayes, Random Forest Classifier*

## I. PENDAHULUAN

Jantung ialah organ yang penting dalam tubuh manusia. Jantung berperan memompa serta mengalirkan darah yang mengangkat oksigen ke segala bagian badan. Penyakit jantung menggambarkan serangkaian keadaan yang mempengaruhi kerja jantung. Sebagian keadaan yang bisa mempengaruhi keadaan kesehatan jantung yakni penyakit pembuluh darah, penyakit arteri koroner, kendala detak( irama) jantung, serta pula cacat jantung bawaan. Sebutan penyakit jantung pula sering berhubungan dengan penyakit kardiovaskular. Kondisi ini biasanya berkaitan dengan kondisi penyempitan atau penyumbatan pembuluh darah, yang dapat menyebabkan serangan jantung, nyeri dada (angina), atau *stroke*. Menurut survei *Sample Registration System* (SRS) pada 2014 Penyakit Jantung Koroner (PJK) menjadi penyebab utama kematian setelah *stroke* pada semua kelompok umur di Indonesia yaitu sebesar 12,9% [1].

Tentunya seiring berkembangnya teknologi, dapat meminimalisir angka kematian karena penyakit jantung. Salah satu teknologi yang ada saat ini adalah *machine learning* yang mampu digunakan untuk mendeteksi penyakit. Misalnya, studi untuk mendiagnosis penyakit hati menggunakan *machine learning* [2] dan studi untuk memprediksi penyakit jantung menggunakan metode *naive bayes* [3]. *Machine learning* sendiri bukanlah teknologi yang menganalisis data dengan akurasi 100% dan menarik kesimpulan berdasarkan analisis data tersebut.

Beberapa model klasifikasi yang bisa digunakan pada *machine learning*, antara lain *k-nearest neighbor (KNN)*, *regresi logistic*, *naive bayes*, *random forest classifier*, dan *support vector machine (SVM)*, *decision tree*. Penelitian tentang klasifikasi penyakit arteri koroner oleh Jefri, Cathlin dan Sentosa. Algoritma yang digunakan pada penelitian tersebut yaitu algoritme C4.5 dan Extreme Learning Machine. Hasilnya menunjukkan bahwa algoritma C4.5 mendapatkan skor akurasi sebesar 93,33%, sedangkan algoritma Extreme Learning Machine mendapatkan skor sebesar 73,33%[4].

Berdasarkan penelitian Jefri, Henry, dan Kenich *Detecting Heart Disease using Machine Learning with Logistic Regression Algorithms*. Hasil penelitian ini menunjukkan algoritma C4.5 lebih baik dari algoritma lainnya, dan akurasi, sensitivitas, presisi dan spesifikasi dari algoritma jaringan saraf lebih rendah dari algoritma lainnya, sedangkan algoritma regresi logistik hanya dalam spesifikasi yang terbaik[5]. Penelitian selanjutnya dari T. Retnasari dan Rahmawati E.yang membandingkan dua algoritma yaitu algoritma *naive bayes* dan C4.5. Hasil penelitian ini menunjukkan bahwa algoritma *naive bayes* mendapatkan skor akurasi sebesar 86,67%, sedangkan algoritma C4.5 mendapatkan skor sebesar 83,70%. [6].

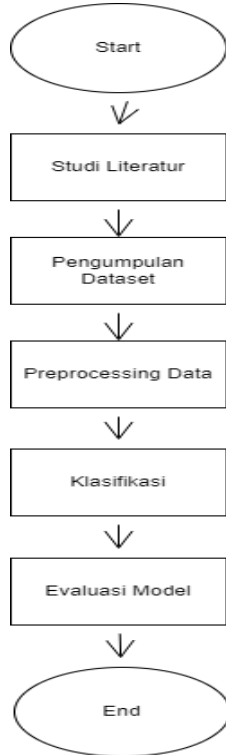
Berdasarkan perbandingan performa algoritma *machine learning* untuk agregasi data social IoT oleh J Meghana, J. Hanumanthapa dan Shiva Prakash, yang membandingkan performa 4 algoritma *decision tree*, *artificial neural network*, *k-nearest neighbors*, dan *naive bayes*. Hasil penelitian menunjukkan bahwa *decision tree* mendapatkan performa yang paling baik dibandingkan algoritma lain [7].

Dengan demikian, penelitian ini peneliti akan membandingkan algoritma *Decision Tree*, *Naive Bayes*, dan *Random Forest Classifier* untuk menemukan algoritma yang paling tepat untuk mengklasifikasikan data penyakit jantung. Parameter yang digunakan sebagai pembanding dari ketiga algoritma tersebut adalah *F1-Score*. Untuk mendapatkan model yang terbaik pada kasus ini saya menggunakan metode *fine*

tuning. Dimana didalam metode *fine tuning* terdapat dua metode yaitu metode *random search* dan *grid search cv*.

## II. METODOLOGI PENELITIAN

Tahapan dalam metodologi penelitian Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung ditunjukkan pada Gbr.1



Gbr. 1 Alur Penelitian

Penelitian ini ada beberapa tahap pada prosesnya supaya penelitian dapat berhasil. Adapun penjelasan dari masing-masing tahapan akan dijelaskan dibawah ini:

### A. Studi Literatur

Tahapan pertama pada penelitian ini adalah tahap studi literatur. Tahap ini dilakukan dengan mencari serta memahami penelitian serupa yang pernah dilakukan sebelumnya untuk mendapatkan teori pendukung serta informasi yang berhubungan dengan penelitian. Literatur yang dipelajari yaitu mengenai penyakit jantung dan metode klasifikasi pada *machine learning*.

### B. Pengumpulan Dataset

Tahap selanjutnya yaitu pengumpulan dataset. Data yang digunakan pada penelitian ini didapatkan dari Kaggle. Data yang digunakan yaitu *Heart Attack Analysis & Prediction Dataset*. Terdapat 12 atribut dan 918 observasi. Variable dalam penelitian ini antara lain.

TABEL I  
VARIABEL PENELITIAN

Atribut	Keterangan	Tipe Data
Age	Umur	Integer
Sex	Jenis kelamin	Object
Chest Pain Type	Tipe nyeri dada	Object
Resting BP	Tekanan darah	Integer
Cholesterol	Kolesterol	Integer
Fasting BS	Gula darah puasa	Integer
Resting ECG	Hasil ECG selama istirahat	Object
Max HR	Detak jantung maksimal	Integer
Exercese Angina	Nyeri dada saat olahraga	Object
Oldpeak	Depresi ST saat olahraga	Float
ST Slope	Kemiringan puncak segmen ST	Object
Heart Disease	Target	Integer

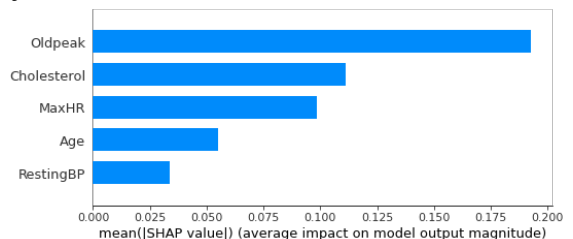
### C. Preprocessing Data

Tahap *preprocessing* merupakan proses dimana data yang sebelumnya tidak terstruktur diubah menjadi data terstruktur untuk diproses lebih lanjut sesuai kebutuhan. Pada tahap *preprocessing* data peneliti melakukan *feature selection*, *handling outlier*, *handling label*. Berikut penjelasan tahapan dari *preprocessing* data.

#### 1) Feature selection

*Feature selection* adalah sesuatu tata cara penganalisaan informasi yang bertujuan guna memilah fitur yang mempengaruhi( fitur maksimal) serta mengesampingkan fitur yang tidak mempengaruhi. Pada sesi *feature selection* periset membagi jadi 2 proses bersumber pada tipe informasi ialah *numeric* serta *categoric*.

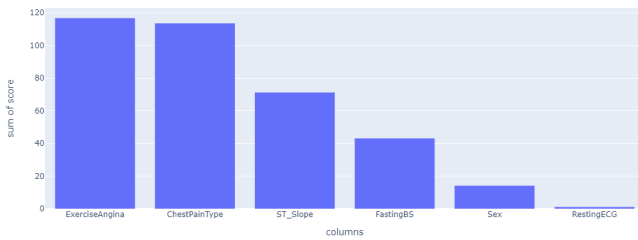
*Feature selection* pada jenis data *numeric* menggunakan model *RandomForestRegressor*. Berikut hasil dari *feature selection* menggunakan model *RandomForestRegressor* ditunjukkan Gbr. 2.



Gbr. 2 Feature selection numeric

Dapat dilihat dari hasil Gbr. 2 bahwa *feature* bertipe *numeric* memiliki relasi terhadap label cukup baik, sehingga *feature* bertipe *numeric* tidak dilakukan *drop feature*.

Selanjutnya *feature selection* pada jenis data *categoric* menggunakan metode *KBest*. Berikut hasil dari *feature selection* menggunakan metode *KBest* ditunjukkan Gbr. 3.

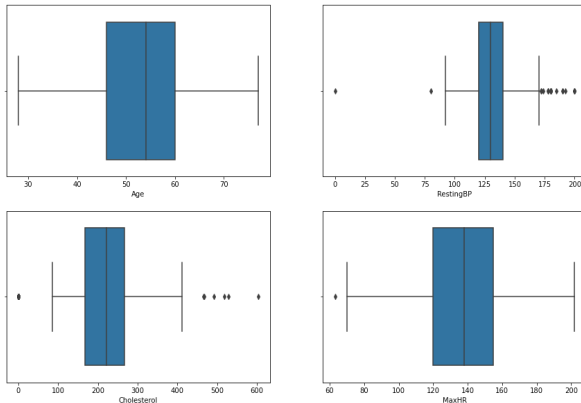


Gbr. 3 Feature selection categoric

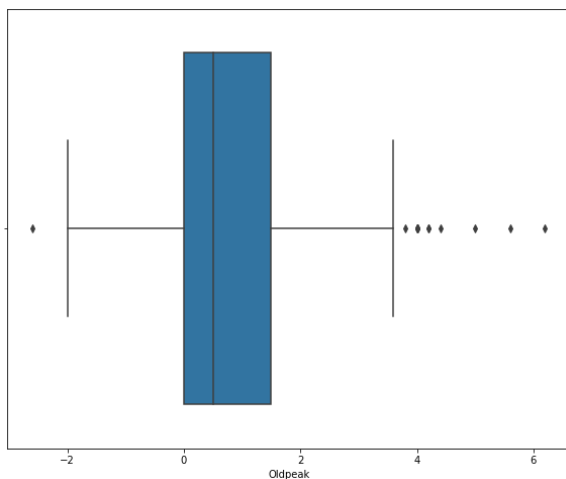
Dapat dilihat dari hasil Gbr. 3 bahwa *feature* bertipe *categoric* yaitu *RestingECG* memiliki nilai relasi yang rendah sehingga dilakukan *drop feature*.

## 2) Handling outlier

Outlier adalah data yang nilainya terlalu jauh dari data lainnya. Adanya perbedaan membuat analisis data terdistorsi atau dengan kata lain tidak mencerminkan fenomena yang sebenarnya. Tahap *handling outlier* dilakukan hanya untuk *feature* bertipe *numeric*. Berikut merupakan grafik *outlier* dari *feature* bertipe *numeric*.



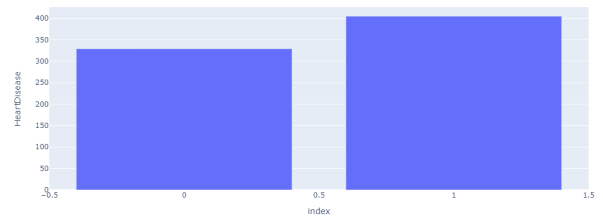
Gbr. 4 Grafik outlier (Age, RestingBP, Cholesterol, MaxHR)



Gbr. 5 Grafik outlier (Oldpeak)

## 3) Handling label

Pada tahap *handling label* peneliti ingin mengetahui kuantiti label pada data training. Berikut merupakan grafik kuantiti label data training.



Gbr. 6 Grafik kuantiti label

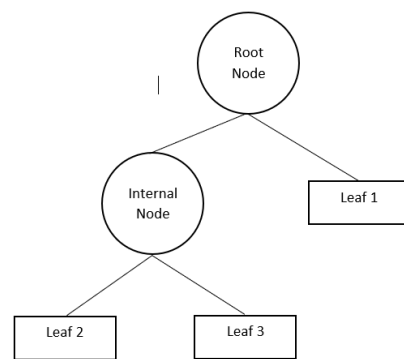
Dapat dilihat pada Gbr. 6 data training memiliki label *balance*.

## D. Klasifikasi

Dalam tahap klasifikasi penelitian ini menggunakan 3 model yakni *Decision Tree*, *Random Forest Classifier*, dan *Naive Bayes*. Penelitian ini akan membandingkan ketiga model tersebut.

### 1) Model Decision Tree

*Decision Tree* ialah salah satu model klasifikasi yang populer, karena mudah bagi orang untuk menginterpretasikannya. *Decision Tree* ialah struktur data yang terdiri dari node(simpul) dan edge. Simpul pohon dibagi menjadi tiga bagian, yakni simpul akar, simpul cabang/internal dan simpul daun. Proses *decision tree* mengubah data tabular menjadi model pohon, kemudian mengubah pohon tersebut menjadi sebuah aturan, kemudian menyederhanakan aturan tersebut. Tujuan model ini adalah memecah proses pengambilan keputusan yang banyak menjadi lebih sederhana, sehingga keputusan yang diperoleh lebih akurat sebagai solusi dari permasalahan.



Gbr. 7 Pohon keputusan

Berdasarkan Gbr. 7 titik akar (root) adalah awal dari sebuah pohon, kemudian setiap cabang pohon merupakan pembagian yang diperoleh dari hasil pengujian, dan titik daun (leaf) merupakan ujung dari pohon yang menggambarkan hasil berupa klasifikasi atau prediksi.

### 2) Model Naïve Bayes

Algoritma *Naive Bayes* ialah salah satu algoritma dari metode klasifikasi. *Naive Bayes* merupakan model klasifikasi berdasarkan metode probabilitas dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes, yang memprediksi kemungkinan masa depan berdasarkan pengalaman masa lampau dengan cara yang dikenal sebagai teori Bayes. Teorema tersebut digabungkan dengan *Naive*, yang mengasumsikan bahwa kondisi antara atribut bervariasi secara bebas. *Naive Bayes* mengasumsikan bahwa ada atau tidak adanya fitur tertentu dari satu kelas tidak ada hubungannya dengan fitur kelas lain.

Berikut merupakan persamaan teorema Bayes :

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

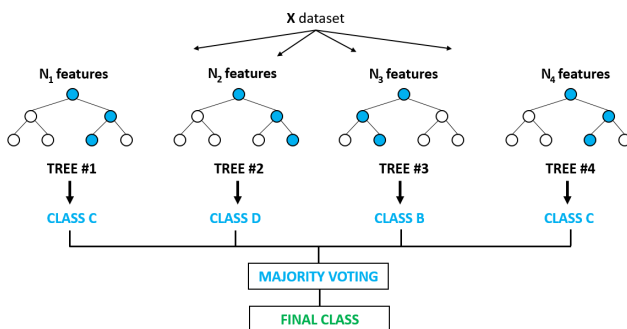
Gbr. 8 Teori Bayes

Keterangan :

- H = Hipotesis data X dari kelas spesifik
- X = Data dengan kelas belum diketahui
- P(X) = Probabilitas X
- P(H|X) = Probabilitas hipotesis H berdasar kondisi X (posterior probability)
- P(H) = Probabilitas hipotesis H (prior probability)
- P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

### 3) Model Random Forest Classifier

*Random Forest (RF)* adalah algoritma untuk mengklasifikasikan jumlah data yang besar. Proses klasifikasi dalam *random forest* dimulai dengan memecah data lalu memasukkan sampel yang ada secara acak ke dalam pohon keputusan. Setelah pohon terbentuk, setiap kelas dicocokkan dengan data sampel lalu dilakukan *voting* kemudian diambil *vote* terbanyak. Berikut diagram alur model *random forest*.



Gbr. 9 Diagram alur *Random Forest*

### E. Evaluasi Model

Tahap evaluasi model dilakukan untuk mendapatkan model yang paling sesuai untuk klasifikasi data penyakit jantung dengan parameter *F1-Score*. Dalam melakukan evaluasi model

ini digunakan metode *fine tuning*, dimana metode *fine tuning* terdapat dua metode yaitu metode *random search* dan *grid search cv* untuk mendapatkan model yang terbaik.

## III. HASIL DAN PEMBAHASAN

Bab ini membahas hasil klasifikasi dari setiap model yang sudah dijalankan dan penjelasan mengenai hasil evaluasi model. Dalam pengujian setiap model, dataset dibagi menjadi dua bagian yaitu 80% data untuk *training* dan 20% data untuk *testing*.

### A. Klasifikasi

Dalam penelitian ini proses klasifikasi setiap model dimulai dengan proses *training* model lalu pengujian terhadap data *training* selanjutnya melakukan *cross validation* untuk mengetahui apakah model mengalami *overfitting* dan melakukan pengujian model terhadap data *dummy*.

Proses *training* setiap model menggunakan data *training* dan setiap model diuji untuk memprediksi hasil akhir dari data *training* tersebut. Dihasilkan nilai akurasi setiap model dari pengujian sebagai berikut.

TABEL II  
HASIL PREDIKSI DATA TRAINING

Model	Akurasi
Decision Tree	1.0 (100 %)
Naïve Bayes	0.87 (87 %)
Random Forest Classifier	1.0 (100 %)

Dapat dilihat pada tabel hasil prediksi menggunakan data *training* diatas. Hasil akurasi pada setiap model mendapatkan nilai yang sangat baik yaitu 100 %.

Selanjutnya peneliti melakukan *cross validation* untuk mengevaluasi kinerja dari setiap model. *Cross validation* merupakan sebuah teknik memvalidasi sebuah model. Teknik ini banyak digunakan untuk membuat prediksi model dan untuk menilai seberapa akurat model memprediksi dalam praktiknya.

Didapatkan hasil *cross validation* dengan parameter akurasi berupa *mean* (rerata) dan standar deviasi. Berikut nilai dari hasil *cross validation* setiap model.

TABEL III  
HASIL CROSS VALIDATION

Model	Akurasi	
	Mean	Standar deviasi
Decision Tree	0.81740	0.01724
Naïve Bayes	0.87602	0.02167
Random Forest Classifier	0.88009	0.01477

Dapat dilihat dari hasil prediksi seetiap model menggunakan data *training* yang ditunjukkan pada tabel I yang mendapatkan hasil akurasi 100 %, peneliti ingin

membuktikan bahwa model tersebut terlalu pintar terhadap data training itu sendiri. Salah satu hal penting dalam mengembangkan suatu model yaitu memastikan apakah model tersebut mengalami *overfitting* atau tidak. *Overfitting* adalah keadaan dimana hasil prediksi terhadap data *training* terlalu baik, namun hasil prediksinya buruk terhadap data *testing*. Untuk mengetahui apakah model tersebut mengalami *overfitting*, peneliti membentuk *range* dari hasil *cross validation* dengan *range* nilai terendah didapatkan dari *mean* – standar deviasi dan nilai tertinggi *mean* + standar deviasi.

TABEL IV  
HASIL RANGE OVERFITTING

Model	Nilai Akurasi	Range	
		Terendah	Tertinggi
Decision Tree	1.0	0.80016	0.83464
Naïve Bayes	0.87	0.85435	0.89769
Random Forest Classifier	1.0	0.86532	0.89486

Dapat dilihat pada tabel III nilai akurasi ketiga model tersebut tidak memenuhi syarat *range* dari hasil *cross validation*. Dapat disimpulkan terdapat dua model mengalami *overfitting*. Untuk membuktikan keakuratan bahwa ketiga model tersebut mengalami *overfitting* atau memiliki kinerja yang buruk, peneliti melakukan pengujian dengan menggunakan dataset *dummy* yaitu diambil dari 10 dataset testing. Didapatkan hasil prediksi sebagai berikut.

TABEL V  
HASIL PREDIKSI DATA TESTING DUMMY

Model	Akurasi
Decision Tree	0.7 (70 %)
Naïve Bayes	0.7 (70 %)
Random Forest Classifier	0.7 (70 %)

Dari hasil prediksi ketiga model tersebut dalam pengujian menggunakan dataset *dummy* menunjukkan bahwa kinerja model belum maksimal.

## B. Evaluasi Model

Pada proses ini setiap model dievaluasi untuk mendapatkan model yang paling cocok digunakan untuk klasifikasi data penyakit jantung. Proses evaluasi model ini menggunakan metode *fine tuning*, dimana *fine tuning* terdapat dua metode yaitu metode *random search* dan *grid search cv*. Pada evaluasi model ini peneliti menggunakan parameter *F1-Score*.

### 1) Evaluasi Model Decision Tree

Proses evaluasi model *decision tree* untuk mencari model terbaik menggunakan metode *random search*, peneliti menggunakan parameter yang terdapat pada model *decision tree*. Berikut parameter model *decision tree* yang digunakan pada penelitian ini.

```
'criterion': ('gini', 'entropy'),
'max_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9],
'min_samples_leaf': [1, 2, 3, 4, 5, 6, 7, 8, 9],
'min_samples_split': [2, 3, 4],
'splitter': ('best', 'random')
```

Pada proses *random search* didapatkan *hyperparameter* berikut.

```
'splitter': 'random',
'min_samples_split': 4,
'min_samples_leaf': 3,
'max_depth': 5,
'criterion': 'entropy'
```

Selanjutnya evaluasi model *decision tree* menggunakan metode *grid search cv* dengan parameter berikut

```
'splitter': ['random'],
'criterion': ['entropy'],
'max_depth': [5],
'min_samples_leaf': [3, 2, 4, 5],
'min_samples_split': [14, 19]
```

Didapatkan *hyperparameter* hasil dari *grid search cv* pada model *decision tree* sebagai berikut.

```
'criterion': 'entropy',
'max_depth': 5,
'min_samples_leaf': 3,
'min_samples_split': 19,
'splitter': 'random'
```

### 2) Evaluasi Model Naïve Bayes

Proses evaluasi model *naïve bayes* untuk mencari model yang terbaik menggunakan metode *random search*, peneliti menggunakan parameter yang terdapat pada model *naïve bayes*. Berikut parameter model *naïve bayes* yang digunakan pada penelitian ini.

```
'alpha': [0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,
3.5, 4.0, 4.5],
'fit_prior': ['True', 'False']
```

Pada proses *random search* didapatkan *hyperparameter* berikut.

```
'fit_prior': 'False', 'alpha': 1.0
```

Selanjutnya evaluasi model *naïve bayes* menggunakan metode *grid search cv* dengan parameter berikut

```
'fit_prior': ['False'],
'alpha': [1.5, 2.0, 2.5, 3.0, 0.5]
```

Pada proses *grid search cv* didapatkan *hyperparameter* sebagai berikut.

```
'alpha': 1.5,
'fit_prior': 'False'
```

### 3) Evaluasi Model Random Forest Classifier

Proses evaluasi model *random forest classifier* untuk mencari model yang terbaik menggunakan metode *random search*, peneliti menggunakan parameter yang terdapat pada model *random forest classifier*. Berikut parameter model *random forest classifier* yang digunakan pada penelitian ini.

```
'n_estimators': [3, 5, 7, 9],
'max_features': ['auto', 'sqrt', 'log2'],
'max_depth': [1, 3, 5, 7, 10],
'min_samples_split': [2, 5, 10],
'min_samples_leaf': [4, 6, 8],
'criterion': ['entropy', 'gini']
```

Pada proses *random search* didapatkan *hyperparameter* berikut.

```
'n_estimators': 9,
'min_samples_split': 10,
'min_samples_leaf': 6,
'max_features': 'sqrt',
'max_depth': 5,
'criterion': 'entropy'
```

Selanjutnya evaluasi model *random forest classifier* menggunakan metode *grid search cv* dengan parameter berikut.

```
'n_estimators': [9, 7, 11, 12, 13],
'min_samples_split': [10, 9, 11],
'max_features': ['sqrt'],
'min_samples_leaf': [6, 5, 7, 8],
'max_depth': [5, 15, 20],
'criterion': ['entropy']
```

Pada proses *grid search cv* didapatkan *hyperparameter* sebagai berikut.

```
'criterion': 'entropy',
'max_depth': 5,
'max_features': 'sqrt',
'min_samples_leaf': 7,
'min_samples_split': 10,
'n_estimators': 9
```

Dari hasil *fine tuning* dengan metode *random search* dan *grid search* pada ketiga model tersebut akan dilakukan pengujian menggunakan dataset testing. Sedangkan parameter nilai performa dari model menggunakan *F1-score*. Berikut hasil pengujian setiap model.

TABEL VI  
HASIL PENGUJIAN DECISION TREE

Pengujian	F1-score		
	Test Tanpa Fine Tuning	Test Random Search	Test Grid Search
1	0.81	0.85	0.84
2	0.83	0.85	0.82
3	0.83	0.84	0.84
4	0.84	0.84	0.86
5	0.83	0.84	0.84
Rata-rata	0.828	0.844	0.84

TABEL VII  
HASIL PENGUJIAN NAÏVE BAYES

Pengujian	F1-score		
	Test Tanpa Fine Tuning	Test Random Search	Test Grid Search
1	0.85	0.85	0.85
2	0.85	0.85	0.85
3	0.85	0.85	0.85
4	0.85	0.85	0.85
5	0.85	0.85	0.85

Rata-rata	0.85	0.85	0.85
-----------	------	------	------

TABEL VIII  
HASIL PENGUJIAN RANDOM FOREST CLASSIFIER

Pengujian	F1-score		
	Test Tanpa Fine Tuning	Test Random Search	Test Grid Search
1	0.86	0.85	0.87
2	0.87	0.87	0.88
3	0.86	0.85	0.87
4	0.85	0.85	0.86
5	0.85	0.84	0.86
Rata-rata	0.858	0.852	0.868

Dapat dilihat rata-rata hasil pengujian setiap model pada tabel VI, tabel VII, tabel VIII bahwa model *random forest classifier* mendapatkan nilai performa *f1-score* tertinggi dibandingkan model *decision tree* dan *naïve bayes* dengan nilai 0.868.

#### KESIMPULAN

Dari hasil pengujian dapat disimpulkan bahwa penerapan model klasifikasi pada data penyakit jantung berhasil. Kemudian dilakukan perbandingan model *decision tree*, *naïve bayes* dan *random forest classifier* setelah dilakukan evaluasi model menggunakan metode *fine tuning* dengan parameter performa *f1-score*.

Dari rata-rata hasil pengujian, model *decision tree* menggunakan evaluasi *random search* mendapatkan nilai 0.844 dan *grid search* 0.84. Kemudian hasil pengujian model *naïve bayes* tidak ada perbedaan antara evaluasi menggunakan *random search* maupun *grid search* yaitu sebesar 0.85. Selanjutnya hasil pengujian model *random forest classifier* menggunakan evaluasi *random search* mendapatkan nilai 0.852 dan *grid search* 0.868. Dengan demikian dapat dilihat dari hasil pengujian bahwa model *random forest classifier* merupakan model terbaik dalam mengklasifikasi penyakit jantung baik itu menggunakan *hyperparameter tuning grid search*.

#### UCAPAN TERIMA KASIH

Puji syukur kehadiran Allah SWT atas rahmat yang telah diberikan, sehingga penelitian ini dapat berjalan dengan lancar tanpa halangan apapun. Serta ucapan terimakasih saya berikan kepada semua pihak yang telah memberikan semangat hingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

- [1] Kementerian Kesehatan RI, "Penyakit jantung penyebab kematian tertinggi, kemenkes ingatkan cerdik," Kementerian Kesehatan RI, 2018. [Daring]. Tersedia: <http://www.padk.kemkes.go.id/health/read/2018/09/26/2/penyakit-jantung-penyebab-kematian-tertinggi-kemenkes-ingatkan-cerdik.html>. [Diakses: 11 Maret 2022].
- [2] P. Elly, S. Qomariyah dan Irhamah, "Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning" INFERENSI, Vol. 2(1), hal. 25 – 32, 2019.
- [3] Bianto, Mufti A., Kusriani dan Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes" Citec Journal, Vol. 6, No. 1, hal. 75 – 83, 2019.
- [4] Pangaribuan, Jefri J., Cathlin T. dan Sentosa W., "Perbandingan Metode Algoritma C4.5 Dan *Extreme Learning Machine* Untuk Mendiagnosis Penyakit Jantung Koroner" INFORMATICS ENGINEERING RESEARCH AND TECHNOLOGY, Vol. 1, No. 1, 2019.
- [5] Pangaribuan, Jefri J., Henry T. dan Kenichi, "Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression" INFORMATION SYSTEM DEVELOPMENT, Vol. 6, No. 2, 2021.
- [6] T. Retnasari dan E. Rahmawati, "Diagnosa prediksi penyakit jantung dengan model algoritma Naïve Bayes dan algoritma C4.5" Konferensi Nasional Ilmu Sosial dan Teknologi (KNiST), hal. 7 – 12, 2017.
- [7] Meghana J, J. Hanumanthappa dan Shiva Prakash, "Performance comparison of machine learning algorithms for data aggregation in social internet of things" Global Transitions Proceedings 2 hal. 212-219, 2021.