

Perbandingan Algoritma *K-Nearest Neighbors* (K-NN) dan *Random forest* terhadap Penyakit Gagal Jantung

Fredilio¹⁾, Julfikar Rahmad²⁾, Stiven Hamonangan Sinurat³⁾, Daniel Ryan Hamonangan Sitompul⁴⁾, Dennis Jusuf Ziegel⁵⁾, Evta Indra^{6)*}

¹⁾²⁾³⁾⁴⁾⁵⁾⁶⁾ Prodi Sistem Informasi, Sains dan Teknologi, Universitas Prima Indonesia

Correspondence author: evtaindra@unprimdn.ac.id, Medan, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1432>

Abstrak

Penelitian ini bertujuan untuk membandingkan akurasi algoritma *K-Nearest Neighbor* (K-NN) dan *Random Forest* dalam mengklasifikasikan penyebab penyakit gagal jantung. Penyakit ini menjadi salah satu penyebab utama kematian di seluruh dunia dan kasusnya terus meningkat di Indonesia. Oleh karena itu, penanganan dan klasifikasi dini terhadap penyebab gagal jantung sangat diperlukan untuk mencegah penyakit tersebut. Penelitian ini diharapkan dapat memberikan informasi tentang metode terbaik untuk mengklasifikasikan penyebab penyakit gagal jantung serta memberikan manfaat bagi tenaga medis dan masyarakat umum dalam menjaga kesehatan jantung mereka.

Kata kunci: Penyakit Gagal Jantung, Klasifikasi, *K-Nearest Neighbors*, *Random Forest*

ABSTRACT

This study aims to compare the accuracy of the K-Nearest Neighbor (K-NN) and Random Forest algorithms in classifying the causes of heart failure. This disease is one of the main causes of death worldwide and its cases continue to increase in Indonesia. Therefore, early treatment and classification of the causes of heart failure are needed to prevent this disease. This research is expected to provide information about the best method for classifying the causes of heart failure and provide benefits for medical personnel and the general public in maintaining their heart health.

Keyword: Heart Failure, Classification, *K-Nearest Neighbor*, *Random Forest*.

PENDAHULUAN

Seiring dengan perkembangan teknologi dibidang kesehatan saat ini, dapat mempermudah tenaga medis dalam menyelesaikan pekerjaannya, baik dalam memberikan pelayanan kepada masyarakat serta dalam melakukan diagnosa penyakit terhadap pasien dengan menggunakan bantuan teknologi (Sutrisno, 2022). Banyak penyakit yang dapat didiagnosis menggunakan bantuan teknologi, yang salah satunya adalah penyakit gagal jantung, penyakit ini merupakan kondisi di mana jantung tidak dapat memompa darah dengan efisien (Desiani et al., 2022). Berdasarkan data statistik dari World Health Organization (WHO) dan Career Development Center (CDC) ada 26,6 juta orang di

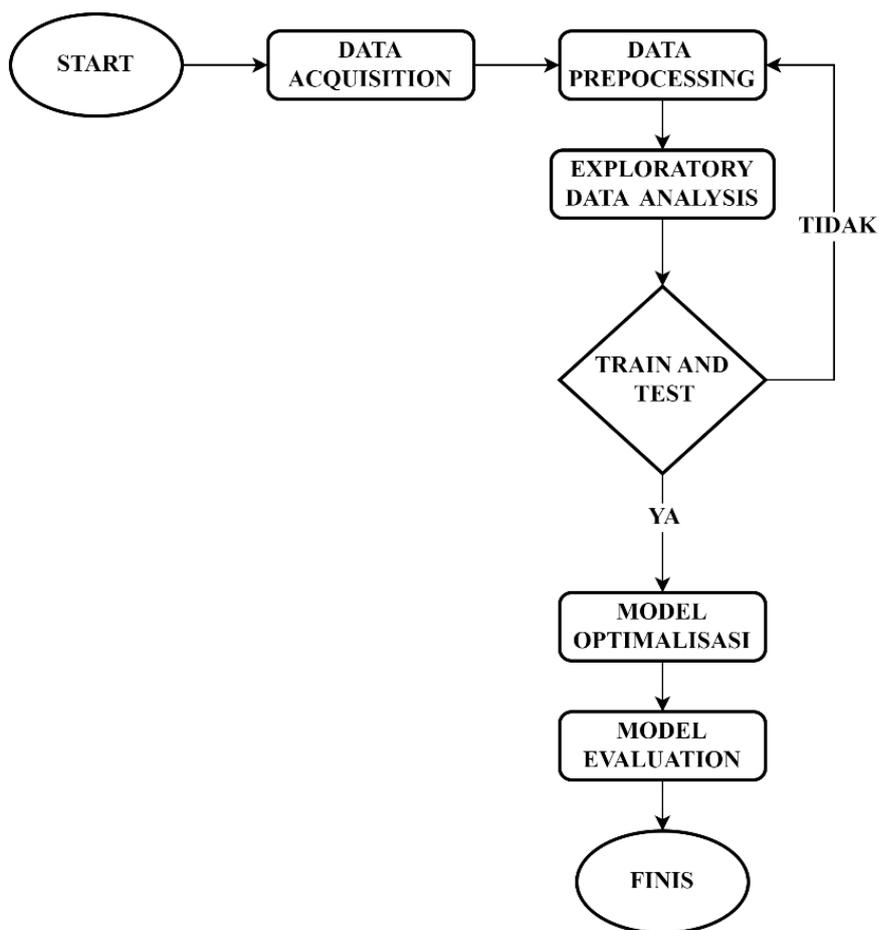
beberapa negara berkembang telah didiagnosis menderita penyakit gagal jantung, menjadikannya salah satu penyebab utama kematian di seluruh dunia dan pada tahun 2020 orang dengan penyakit gagal jantung di Indonesia meningkat hampir 24% lebih banyak dibandingkan tahun 2005 atau setara dengan 82 juta jiwa penderita gagal jantung (Disease, 2022). Jantung merupakan organ vital yang berfungsi untuk memompa dan mengalirkan darah ke seluruh tubuh, sehingga organ dan jaringan tubuh dapat menjalankan fungsinya dengan baik namun, ada beberapa hal yang bisa menyebabkan gangguan masalah pada jantung yang membuat organ ini tidak bisa berfungsi secara normal, ataupun biasa disebut dengan Gagal Jantung (Lumi et al., 2021). Permasalahan yang sering dihadapi oleh pasien ataupun orang awam adalah tidak dapat mengetahui secara pasti penyebab dan faktor penyebab dari penyakit gagal jantung sehingga mereka tidak dapat menjegah penyakit tersebut, serta menentukan seberapa besar akurasi dari penyebab gagal jantung, dengan adanya permasalahan tersebut perlu adanya klasifikasi dari penyebab penyakit gagal jantung sehingga dapat meminimalisir terkena penyakit tersebut dan dapat menghindarinya.

Penelitian-penelitian sebelumnya yang membahas tentang penyakit gagal jantung, sebagai referensi peneliti yaitu: “Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung”(Hasanah et al., 2022), penelitian ini membahas mengenai implementasi algoritma naïve bayes dalam mengklasifikasi penyakit gagal jantung dengan jumlah pasien sebanyak 100 orang, selanjutnya “Kualitas Hidup Penderita Gagal Jantung Kongestif Berdasarkan Derajat Kemampuan Fisik dan Durasi Penyakit”(Saida et al., 2020), penelitian tersebut membahas tentang menentukan faktor-faktor yang berhubungan dengan kualitas hidup penderita gagal jantung kongestif berdasarkan derajat kemampuan fisik dan durasi penyakit, “Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular”(Putri, 2021), penelitian ini membahas pengukuran performa (akurasi, presisi, recall dan f-measure) metode knn dan crossvalidation pada penyakit cardiovascular. Mengacu pada referensi yang telah dipaparkan, maka melakukan penanganan sejak dini dan melakukan klasifikasi penyebab penyakit gagal jantung sangatlah diperlukan dalam mencegah penyakit gagal jantung. Berdasarkan latar belakang, maka penelitian ini berfokus pada “Perbandingan Algoritma K-Nearest Neighbor(K-NN) Dan Randomforest Terhadap Penyakit Gagal Jantung”

METODE

Kemajuan teknologi dibidang kesehatan tentu berdampak besar pada proses mendiagnosis penyakit serta mengklasifikasi penyebab dari penyakit gagal jantung, dengan adanya kemajuan tersebut semua hal dapat lebih cepat menemukan hasil diagnosis nya dengan bantuan machine learning (Prihatiningsih & Sudyasih, 2018). *Machine learning* adalah kecerdasan buatan (*Artificial intelligence*) yang memungkinkan komputer untuk belajar dari data telah dibuat daripada harus mengikuti instruksi yang telah ditentukan sebelumnya secara eksplisit(perintah langsung)(Wahyono, 2018). Pada penelitian ini menggunakan bantuan kecerdasan buatan (*Artificial intelligence*) untuk mengklasifi faktor-faktor penyebab dari penyakit gagal jantung dengan algoritma *K-Nearest Neighbor*(K-NN) dan *Randomforest*. Algoritma *K-Nearest Neighbor*(K-NN) adalah pendekatan klasifikasi yang menetapkan kategori berdasarkan mayoritas kategori, menetapkan kesamaan antara dua vektor, dan menetapkan kesamaan antara kumpulan data (Reza Noviansyah et al., 2018). Algoritma *Random forest* adalah Sebuah metode atau algoritma dari pendekatan pohon keputusan dari turunan metode CART (*Classification and Regression Trees*) (Adrian et al., 2021).

Agar penelitian ini dapat berjalan dengan baik dan selesai dengan tepat waktu maka dibuat prosedur kerja. Adapun prosedur kerja penelitian ini adalah sebagai berikut:



Gambar 1. Flowchat

Data acquisition merupakan proses yang dilakukan oleh peneliti untuk mengumpulkan data, dalam penelitian ini data yang digunakan dataset pasien penyakit gagal jantung dengan jumlah data sebanyak 299 baris dan 13 kolom (Anggara et al., 2022). Seperti dilihat pada gambar 2 berikut:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets
0	75.0	0	582	0	20	1	265000.00
1	55.0	0	7861	0	38	0	263358.03
2	65.0	0	146	0	20	0	162000.00
3	50.0	1	111	0	20	0	210000.00
4	65.0	1	160	1	20	0	327000.00
...
294	62.0	0	61	1	38	1	155000.00
295	55.0	0	1820	0	38	0	270000.00
296	45.0	0	2060	1	60	0	742000.00
297	45.0	0	2413	0	38	0	140000.00
298	50.0	0	196	0	45	0	395000.00

299 rows x 13 columns

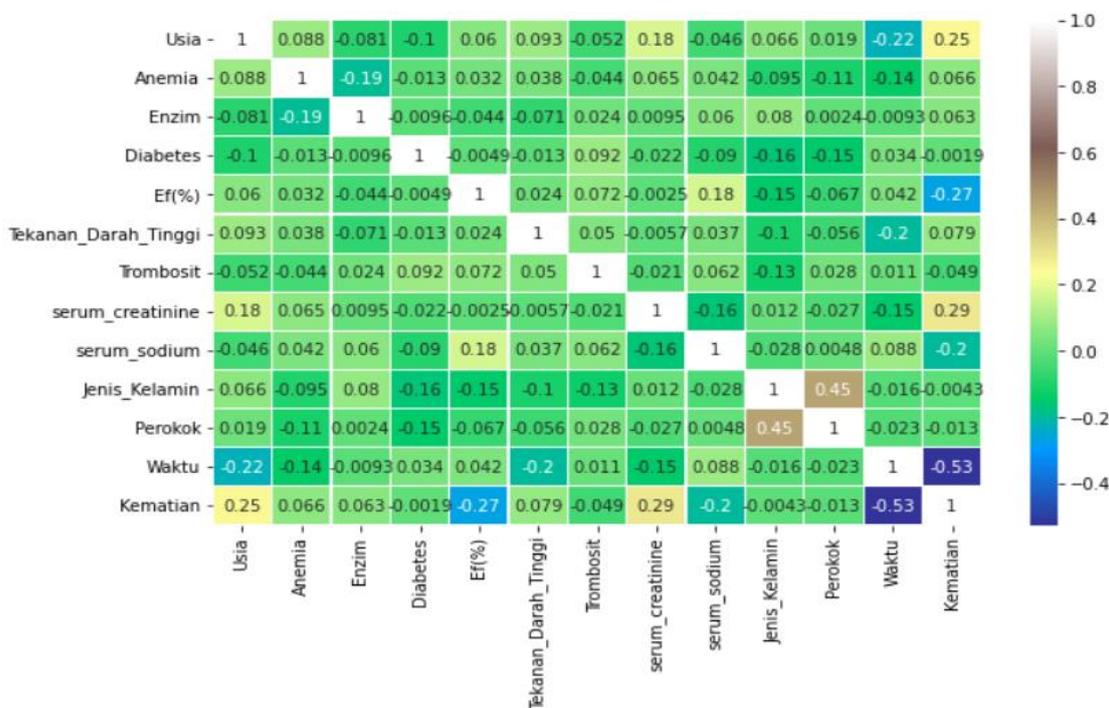
Gambar 2. Dataset pasien gagal jantung

Data preprocessing adalah proses mengolah data mentah menjadi format data yang lebih efisien sehingga model yang dibuat menjadi lebih bagus, hal tersebut bertujuan untuk mendapatkan hasil yang akurat (Said et al., 2022). Seperti dapat dilihat pada gambar 3 berikut:

```
#mengganti nama kolom menjadi bahasa indonesia
df.rename(columns={'age': 'Usia',
                  'anaemia': 'Anemia',
                  'creatinine_phosphokinase': 'Enzim',
                  'diabetes': 'Diabetes',
                  'ejection_fraction': 'Ef(%)',
                  'high_blood_pressure': 'Tekanan_Darah_Tinggi',
                  'platelets': 'Trombosit',
                  'sex': 'Jenis_Kelamin',
                  'smoking': 'Perokok',
                  'time': 'Waktu',
                  'DEATH_EVENT': 'Kematian'}, inplace=True)
```

Gambar 3. Data Preprocessing

Exploratory Data Analysis (EDA) adalah proses eksplorasi data yang bertujuan untuk memahami isi dan komponen penyusun data. Pada penelitian ini, EDA dilakukan pada dataset untuk melihat isi data, korelasi data, dan persebaran data. Seperti dapat dilihat pada gambar 4 berikut:



Gambar 4. Exploratory Data Analysis

Train and test adalah proses dimana dimana dataset di olah dengan menggunakan algoritma *K-Nearest Neighbors* (K-NN) Dan *Randomforest* untuk melakukan training data yang bertujuan untuk melihat keakuratan model yang telah dibuat (Amos, 2013). Seperti dilihat pada gambar 5 berikut:

```
#import library random foresd
from sklearn.ensemble import RandomForestClassifier

rf= RandomForestClassifier()
rf.fit(X_train,Y_train)

RandomForestClassifier()

Y_pred_rf = rf.predict(X_test)

accuracy_rf=accuracy_score(Y_test,Y_pred_rf)*100
accuracy_rf

99.5
```

Gambar 5. Proses *Train And Test*

Model optimalisasi adalah suatu proses meningkatkan presisi model yang telah ditraining menggunakan algoritma *K-Nearest Neighbor*(K-NN) dan *Randomforest* bertujuan untuk mendapatkan model yang lebih optimal (Prima et al., 2021). Proses tersebut dapat dilihat pada gambar 6 berikut:

```
# IMPORT LIBRARY GRIDSEARCHCV UNTUK HYPER-PARAMETER TUNING
from sklearn.model_selection import GridSearchCV

# DEKLARASI PARAMETER YANG AKAN DIPAKAI BESERTA VALUENYA
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'kernel': ['rbf']}
```

Gambar 6. Proses optimalisasi model

Model evaluation merupakan proses mengidentifikasi model yang bertujuan untuk mengoptimalisasi model yang telah ditraining dan testing agar mendapatkan nilai akurasi yang bagus dan untuk menemukan model terbaik yang mewakili data kita (Moriasi et al., 2007). Seperti dapat dilihat padagambar 7 berikut:

```
#MODEL EVALUATION
# IMPORT LIBRARY EVALUATION
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.metrics import roc_curve, roc_auc_score
from sklearn.metrics import classification_report
```

Gambar 7. Proses *Model Evaluation*

HASIL DAN PEMBAHASAN

Pada penelitian ini dataset yang berbentuk bahasa Inggris diubah kedalam bahasa Indonesia, seperti 'age', 'Usia', 'anaemia', 'Anemia', 'creatinine_phosphokinase', 'Enzim', 'diabetes', 'Diabetes', 'ejection_fraction', 'Ef(%)', 'high_blood_pressure', 'Tekanan Darah Tinggi', 'platelets', 'Trombosit', 'sex', 'Jenis_Kelamin', 'smoking', 'Perokok', 'time', 'Waktu', 'Death_Event', 'Kematian' proses tersebut dilakukan bertujuan untuk mempermudah membaca dataset. Seperti dilihat pada gambar 8 berikut.

```
#mengganti nama kolom menjadi bahasa indonesia
df.rename(columns={'age': 'Usia',
                  'anaemia': 'Anemia',
                  'creatinine_phosphokinase': 'Enzim',
                  'diabetes': 'Diabetes',
                  'ejection_fraction': 'Ef(%)',
                  'high_blood_pressure': 'Tekanan_Darah_Tinggi',
                  'platelets': 'Trombosit',
                  'sex': 'Jenis_Kelamin',
                  'smoking': 'Perokok',
                  'time': 'Waktu',
                  'DEATH_EVENT': 'Kematian'}, inplace=True)
```

Gambar 8. Proses *Rename Columns*

Proses menghapus *duplicate* data dengan menggunakan *syntax* `df=df.drop_duplicates()` bertujuan untuk membersihkan data yang sama sehingga tidak mempengaruhi model pada saat training dan testing. Seperti dapat dilihat pada gambar 9 berikut.

```
#Menghapus data yang di duplikasi
df=df.drop_duplicates()

df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 299 entries, 0 to 298
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Usia                   299 non-null    int64
1   Anemia                 299 non-null    int64
2   Enzim                  299 non-null    int64
3   Diabetes               299 non-null    int64
4   Ef(%)                  299 non-null    int64
5   Tekanan_Darah_Tinggi  299 non-null    int64
6   Trombosit              299 non-null    int64
7   serum_creatinine      299 non-null    int64
8   serum_sodium           299 non-null    int64
9   Jenis_Kelamin         299 non-null    int64
10  Perokok                299 non-null    int64
11  Waktu                  299 non-null    int64
12  Kematian               299 non-null    int64
dtypes: int64(13)
memory usage: 32.7 KB
```

Gambar 9. Proses *Delete Duplicate Data*

Proses melihat kelas ini bertujuan untuk melihat jumlah kelas ‘kematian’ berapakah yang bernilai 0 dan 1 yang dimana 0 merupakan jumlah pasien yang mati sedangkan 1 adalah jumlah pasien tidak mati, pada penelitian ini jumlah kematian sebanyak 203 sedangkan yang tidak mati sebanyak 96, tahapan tersebut dapat dilakukan dengan menggunakan *syntax* print (df['Kematian'].value_counts ()) cls_0=df[df['Kematian'] ==0] cls_1=df[df['Kematian'] ==1]. Dapat dilihat pada gambar 10 berikut.

```
#Melihat jumlah kelas 0 dan 1
print(df['Kematian'].value_counts())
cls_0=df[df['Kematian']==0]
cls_1=df[df['Kematian']==1]

0    203
1     96
```

Gambar 10. Proses *View the number of classes 0 and 1*

Pengambilan sampel 500, replacee strategi statistik untuk menemukan pola dan tren dalam kumpulan data yang lebih luas dengan memilih, memanipulasi, dan menganalisis sampel titik data yang representatif, pada penelitian ini pengambilan data sampling dilakukan pada kelas ‘kematian’ dengan jumlah sampling sebanyak 500 data. Dapat dilihat pada gambar 11 berikut.

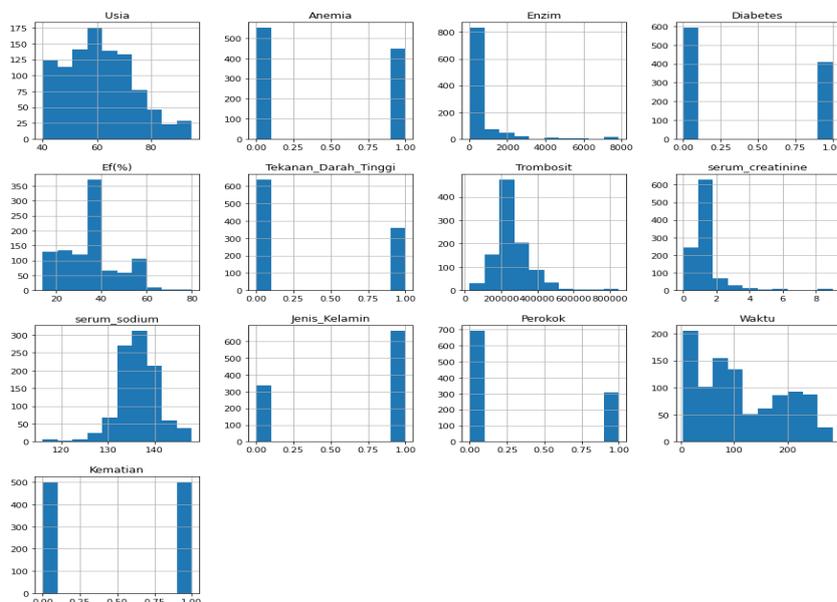
```
#Melihat jumlah kelas 0 dan 1(sampling data)

cls_0=cls_0.sample(500,replace=True)
cls_1=cls_1.sample(500,replace=True)
df=pd.concat([cls_0,cls_1],axis=0)
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 174 to 67
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Usia                   1000 non-null   int64
1   Anemia                 1000 non-null   int64
2   Enzim                  1000 non-null   int64
3   Diabetes               1000 non-null   int64
4   Ef(%)                  1000 non-null   int64
5   Tekanan_Darah_Tinggi  1000 non-null   int64
6   Trombosit              1000 non-null   int64
7   serum_creatinine      1000 non-null   int64
8   serum_sodium           1000 non-null   int64
9   Jenis_Kelamin         1000 non-null   int64
10  Perokok                1000 non-null   int64
11  Waktu                  1000 non-null   int64
12  Kematian               1000 non-null   int64
dtypes: int64(13)
memory usage: 109.4 KB
```

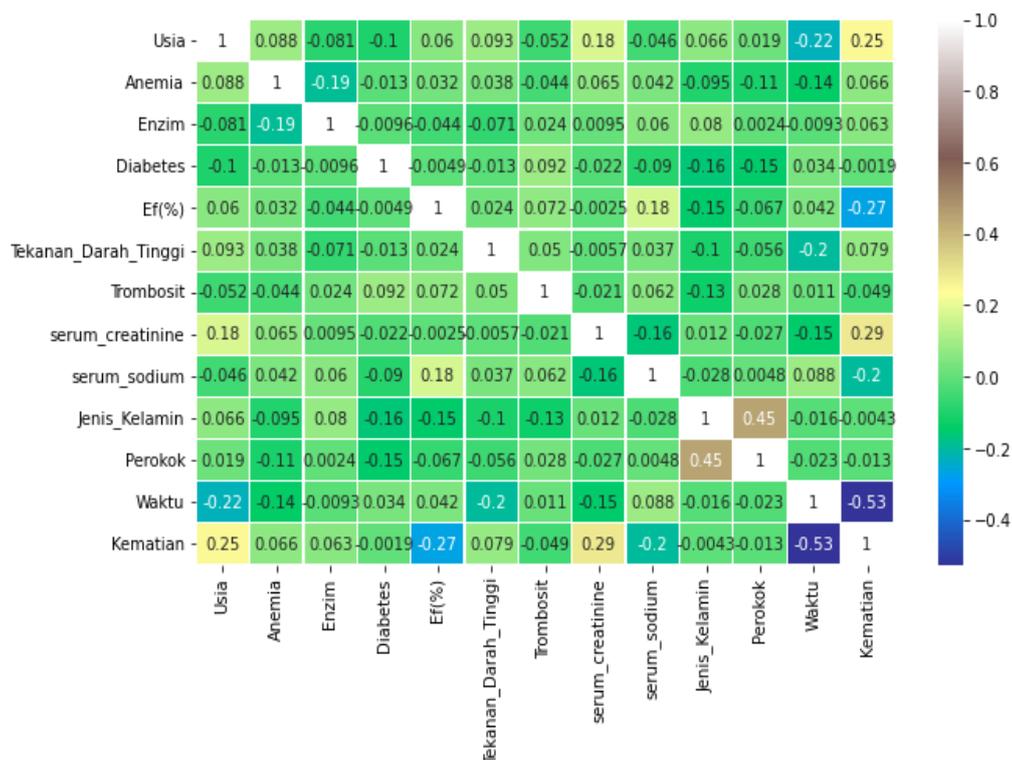
Gambar 11. Proses *Sampling Data*

Pada gambar 12 ini menyajikan penyebaran data pada dataset dengan menggunakan histogram yang bertujuan untuk memudahkan dalam membaca data pasien penyakit gagal jantung, data tersebut berupa diagram usia, enzim, ef(%), trombosit, dan waktu. Pada histogram terlihat pasien dengan usia 40 tahun sebanyak 125 orang, pasien dengan umur 60 tahun sebanyak 175 orang dan pasien dengan usia 80 tahun sebanyak 49 orang, seperti dapat dilihat pada gambar 12 berikut.



Gambar 12. Visualisasi penyebaran dataset gagal jantung

Membuat tabel korelasi bertujuan untuk melihat hubungan antar variabel yang saling berhubungan dengan ketentuan semakin mendekati angka 1 semakin berhubungan variabel tersebut, seperti dapat dilihat pada gambar 13 variabel yang saling berhubungan atau berkaitan itu ada variabel perokok dengan variabel jenis kelamin.

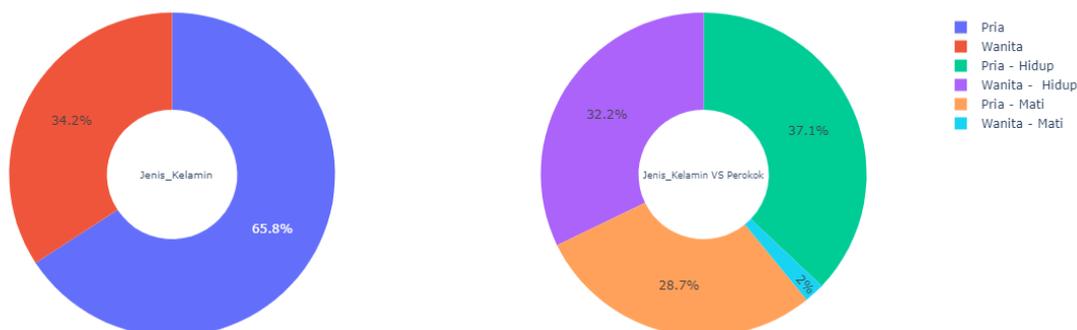


Gambar 13. Tabel Korelasi Penyakit Gagal Jantung

Membuat distribusi jenis kelamin bertujuan untuk memudahkan membaca dan mengetahui informasi pada dataset. Dapat dilihat pada diagram lingkaran jenis kelamin menunjukkan jenis kelamin pria lebih banyak dengan jumlah 65,8% sedangkan wanita sebanyak 34,2% , jumlah tersebut menunjukkan bahwa pada dataset lebih banyak jenis kelamin pria dari pada wanita, serta dapat dilihat juga pada diagram lingkaran jenis kelamin vs perokok, pria hidup karena perokok sebanyak 37,1%, pria mati karena perokok sebanyak 28,7%, untuk wanita mati karena perokok sebanyak 2%, dan wanita hidup karena perokok sebanyak 32,2%, sehingga dapat disimpulkan bahwa lebih banyak pria mati karena perokok dari pada wanita mati karena merokok yang dapat dilihat pada gambar 14 berikut.

GENDER DISTRIBUTION IN THE DATASET

Jenis_Kelamin VS Perokok



Gambar 14. Gender Distribution in The Dataset

Train and Test dataset merupakan tahapan memisahkan variabel X dan Y menjadi dua bagian, variabel X meliputi Usia, Anemia, Enzim, Diabetes, Ef (%), Tekanan Darah Tinggi, Trombosit, serum creatinine, serum sodium, Jenis Kelamin, Perokok, sedangkan variabel Y meliputi kematian dan waktu ini dilakukan untuk melakukan *train and test*. *Train and Test* dataset dilakukan dengan menggunakan library *python scikit-learn*. Hasil pembagian train and test 80:20 dengan random state 101 yang dapat dilihat pada gambar 15.

```
#import library data partitioning/splitting data dan konfigurasi
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test= train_test_split
(X,Y, test_size=0.2, random_state=101,shuffle=True)
```

Gambar 15. Train and Test Splitting dataset

Hasil pembuatan model train dan test algoritma *K-Nearest Neighbor*(K-NN) memperoleh hasil accuracy score sebesar 86%, *accuracy on training set* sebesar 93%, *accuracy on test set* sebesar 86%, seperti yang dapat dilihat pada gambar 16 berikut.

```
accuracy_knn = accuracy_score(Y_test,y_pred_knn)*100
accuracy_knn

86.0

print("Accuracy on training set: {:.3f}".format(knn.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(knn.score(X_test, Y_test)))

Accuracy on training set: 0.936
Accuracy on test set: 0.860
```

Gambar 16. Hasil train and test algoritma *K-Nearest Neighbor* (K-NN)

Hasil pembuatan model train dan test algoritma *Randomforest* memperoleh hasil *accuracy score* sebesar 96%, *accuracy on training set* sebesar 98%, *accuracy on test set* sebesar 96%, seperti yang dapat dilihat pada gambar 17 berikut.

```
accuracy_grid_rf = accuracy_score(Y_test,y_pred_grid_rf)*100
accuracy_grid_rf

96.5

print("Accuracy on training set: {:.3f}".format(grid_rf.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(grid_rf.score(X_test, Y_test)))

Accuracy on training set: 0.988
Accuracy on test set: 0.965
```

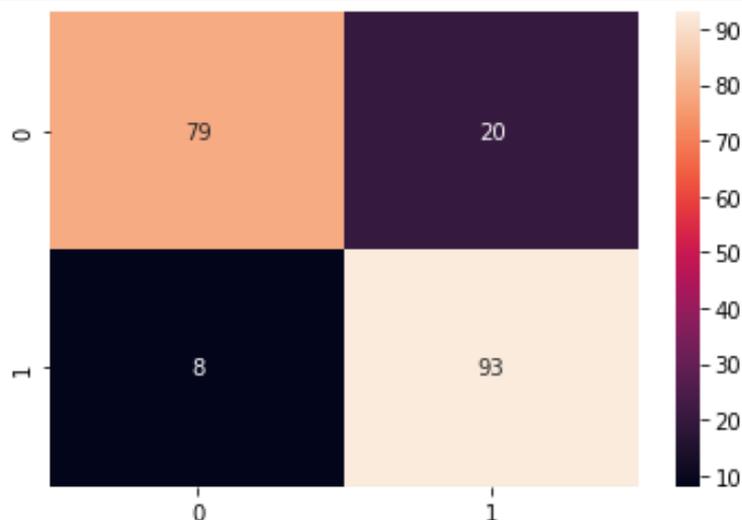
Gambar 17. Hasil Train And Test Algoritma *Randomforest*

Pembuatan model evaluations merupakan proses akhir dimana proses ini menggunakan *Confusion Matrix* untuk menampilkan data pasien gagal jantung, dengan mengimport library `from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, from sklearn.metrics import roc_curve, roc_auc_score`, dan `from sklearn.metrics import classification_report` sebagai syntax untuk menampilkan hasil gambar *Confusion Matrix*, *classification report*, seperti yang dapat dilihat pada gambar 18.

```
#MODEL EVALUATION
# IMPORT LIBRARY EVALUATION
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.metrics import roc_curve, roc_auc_score
from sklearn.metrics import classification_report
```

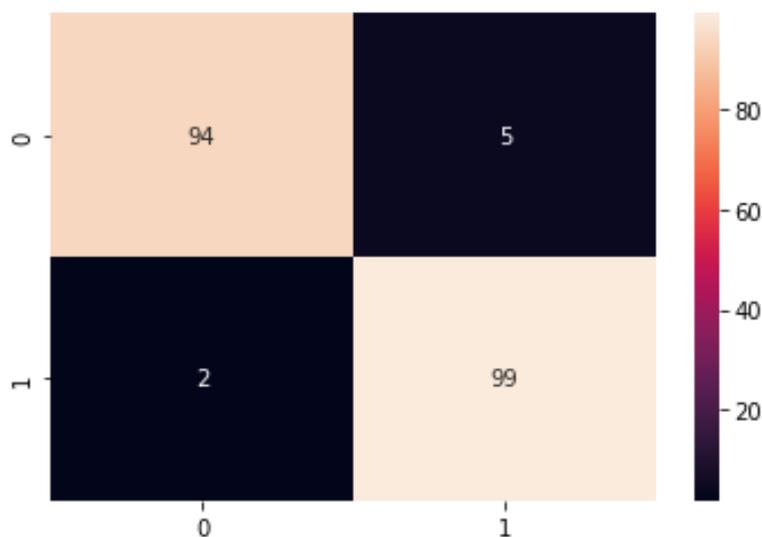
Gambar 18. Import Library Evaluation

Hasil dari model evaluation algoritma *K-Nearest Neighbor*(K-NN) pada bagian kiri atas model dapat menebak kelas 0 sebanyak 79 kali benar dari 99 kali percobaan dan pada bagian kanan bawah model dapat menebak kelas 1 sebanyak 93 kali benar dari 101 percobaan, seperti dapat dilihat pada gambar 19 berikut.



Gambar 19. *Confusion Matrix* Algoritma *K-Nearest Neighbor*(K-NN)

Hasil dari model evaluation algoritma *Random forest* pada bagian kiri atas model dapat menebak kelas 0 sebanyak 94 kali benar dari 99 kali percobaan dan pada bagian kanan bawah model dapat menebak kelas 1 sebanyak 99 kali benar dari 101 percobaan, seperti dapat dilihat pada gambar 20 berikut.



Gambar 20. *Confusion Matrix* Algoritma *Random Forest*

KESIMPULAN DAN REKOMENDASI

Dengan adanya kemajuan teknologi pada bidang kesehatan diharapkan dapat membantu pasien dalam mengetahui faktor-faktor apa saja yang menjadi penyebab penyakit gagal jantung dengan menggunakan bantuan machine learning berbasis algoritma klasifikasi, sehingga dengan parameter penyakit gagal jantung dapat diprediksi sejak dini oleh pasien. Proses-proses yang dilakukan meliputi *rename columns*, menghapus data duplikat, melihat jumlah kelas 0 dan 1, dan sampling data. Analisis data eksploratif dilakukan dengan memeriksa penyebaran data, membuat tabel korelasi, membuat distribusi jenis kelamin, dan membuat diagram box plot. Dari analisis, terlihat bahwa usia 40 tahun memiliki 125 pasien, usia 60 tahun memiliki 175 pasien, dan usia 80 tahun memiliki 49 pasien. Variabel perokok memiliki hubungan dengan variabel jenis kelamin. Model *machine learning* yang telah dibuat dapat menentukan faktor apa yang menyebabkan penyakit gagal jantung pada pasien. Berdasarkan hasil dari analisis dua algoritma *K-Nearest Neighbor* (K-NN) dan algoritma *Random forest*, maka diperoleh hasil akurasi terbaik dari algoritma *Random Forest* sebesar 96,5 %

Direkomendasikan kedepannya terus melakukan penelitian dan pengembangan teknologi dalam bidang kesehatan untuk mempermudah tenaga medis dalam memberikan pelayanan dan diagnosa penyakit, termasuk penyakit gagal jantung. Menyelesaikan masalah ketidaktahuan masyarakat dan pasien terhadap penyebab dan faktor penyebab penyakit gagal jantung melalui klasifikasi dan edukasi. Fokus pada penanganan dini penyakit gagal jantung dengan membandingkan dan mengevaluasi performa dari algoritma *K-Nearest Neighbor* (K-NN) dan *Randomforest* dalam mengatasi masalah tersebut.

REFERENSI

- Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest Dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*, 7(1), 36–40. <https://doi.org/10.26877/jiu.v7i1.7099>
- Amos, S. (2013). When Training And Test Sets Are Different: Characterizing Learning Transfer. *Dataset Shift In Machine Learning*, 2–28. <https://doi.org/10.7551/mitpress/9780262170055.003.0001>

- Anggara, Z. T., Dzulqarnain, M. F., & Sujatmiko, W. (2022). *Visualisasi Data Citra Untuk Klasifikasi Kalimantan ' S Batik Production Menggunakan Neural Network Website-Based Gymship Management System Backend Design At Alterra Academy*. 1(2), 32–37.
- Desiani, A., Akbar, M., Irmeilyana, & Amran, A. (2022). Implementasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. *Jurnal Teknik Elektro Dan Komputasi (ELKOM)*, 4(2), 207–214.
- Disease, H. F. (2022). *Penerapan Algoritma Decision Tree C5 . 0 Untuk Memprediksi Tingkat Kematian Pasien Penyakit Gagal Jantung*. 4(02), 216–222.
- Hasanah, Q., Oktavianto, H., & Rahayu, Y. D. (2022). Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung. *Jurnal Smart Teknologi*, 3(4), 382–389.
- Lumi, A. P., Joseph, V. F. F., & Polii, N. C. I. (2021). Rehabilitasi Jantung Pada Pasien Gagal Jantung Kronik. *Jurnal Biomedik:JBM*, 13(3), 309. <https://doi.org/10.35790/Jbm.V13i3.33448>
- Moriassi, D. N., Arnold, J. G., Liew, M. W. Van, Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). *Model Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations*. 50(3), 885–900.
- Prihatiningsih, D., & Sudyasih, T. (2018). Perawatan Diri Pada Pasien Gagal Jantung. *Jurnal Pendidikan Keperawatan Indonesia*, 4(2). <https://doi.org/10.17509/Jpki.V4i2.13443>
- Prima, J., Sistem, J., Komputer, I., No, V., Radhi, M., Ryan, D., Sitompul, H., Sinurat, S. H., & Indra, E. (2021). *Prediksi Harga Mobil Menggunakan Algoritma Regressi Dengan Hyper-Parameter Tuning*. 4(2), 1–5.
- Putri, I. P. (2021). Analisis Performa Metode K- Nearest Neighbor (KNN) Dan Crossvalidation Pada Data Penyakit Cardiovascular. *Indonesian Journal Of Data And Science*, 2(1), 21–28. <https://doi.org/10.33096/Ijodas.V2i1.25>
- Reza Noviansyah, M., Rismawan, T., Marisa Midyanti, D., (2018). Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data AWS (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48–56.
- Said, H., Matondang, N., Nurramdhani Irmanda, H., & Informasi, S. (2022). *Penerapan*

Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi Application Of K-Nearest Neighbor Algorithm To Predict Consumable Water Quality. 21(2), 256–267. Wwww.Kaggle.Com

Saida, S., Haryati, H., & Rangki, L. (2020). Kualitas Hidup Penderita Gagal Jantung Kongestif Berdasarkan Derajat Kemampuan Fisik Dan Durasi Penyakit. *Faletahan Health Journal*, 7(02), 70–76. <https://doi.org/10.33746/Fhj.V7i02.134>

Sutrisno, S. (2022). Sistem Pakar Mendiagnosa Penyakit Gagal Jantung Pada Manusia Dengan Menggunakan Metode Certainty Factor Berbasis Web. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 5(1), 20–27. <https://doi.org/10.55338/Jikoms.V5i1.207>.

Wahyono, T. (2018). Fundamental Of Python For Machine Learning: Dasar-Dasar Pemrograman Python Untuk Machine Learning Dan Kecerdasan Buatan. *Gava Media*, September 2018, 49.