

PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST UNTUK PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE FEATURE SELECTION ANOVA F VALUE

PENULIS

¹⁾Andini Putri Humaira, ²⁾Sari Ningsih

ABSTRAK

Penyakit jantung masih menjadi salah satu penyebab kematian tertinggi secara global dengan angka kejadian yang terus meningkat. Penelitian ini membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest (RF) dalam memprediksi risiko penyakit jantung menggunakan dataset dari Kaggle. Untuk meningkatkan performa model, diterapkan metode seleksi fitur ANOVA F-value guna mengidentifikasi atribut yang paling relevan. Hasil pengujian menunjukkan bahwa Random Forest memiliki performa lebih baik dengan akurasi 99%, dibandingkan SVM yang mencapai 97%. Berdasarkan evaluasi confusion matrix, Random Forest juga menghasilkan tingkat kesalahan klasifikasi yang lebih rendah. Oleh karena itu, Random Forest dinilai lebih efektif untuk prediksi penyakit jantung dan berpotensi menjadi dasar pengembangan aplikasi deteksi dini berbasis machine learning. Penelitian ini diharapkan dapat mendukung pemanfaatan kecerdasan buatan dalam meningkatkan upaya pencegahan dan deteksi dini penyakit jantung.

Kata Kunci

Penyakit Jantung, Machine Learning, SVM, Random Forest, ANOVA F-value, Feature Selection.

AFILIASI

Prodi, Fakultas

Nama Institusi

Alamat Institusi

¹⁾Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika

²⁾Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika

^{1,2)}Universitas Nasional

^{1,2)}Jl.Sawo Manila No.61, RT.14/RW.7, Pejaten Bar.,Ps.Minggu, Jakarta Selatan, Daerah Khusus Ibukota Jakarta

KORESPONDENSI

Penulis

Email

Sari Ningsih

lectures.sariningsih@gmail.com

LICENSE



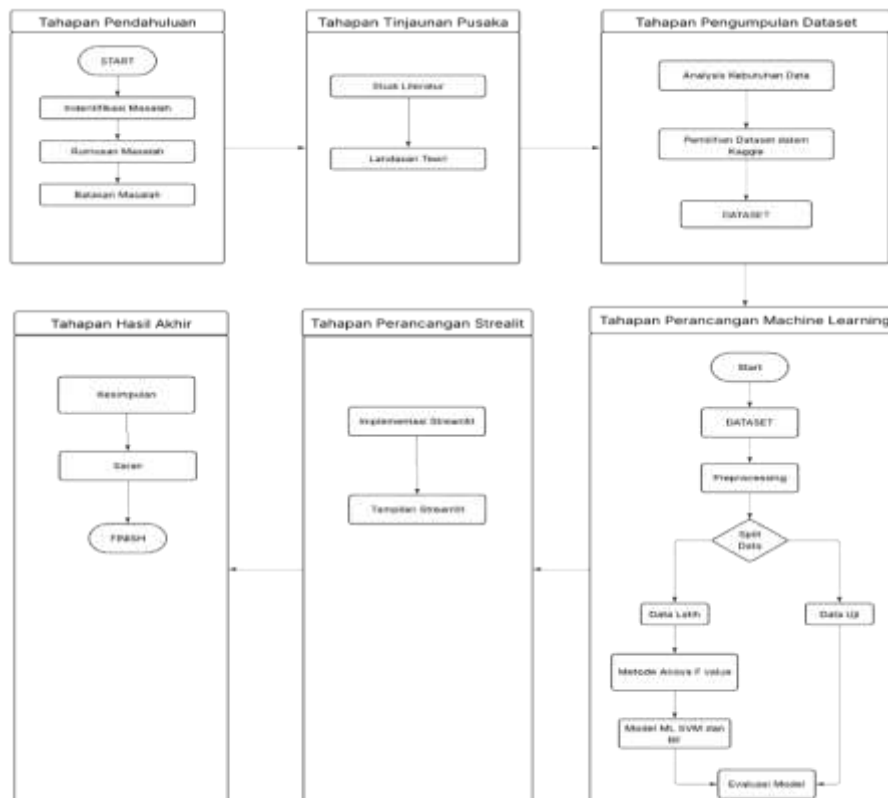
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

Infark miokard, atau yang lebih dikenal sebagai serangan jantung, merupakan salah satu bentuk gangguan jantung yang tergolong sangat berbahaya. Proses identifikasi penyakit ini biasanya melibatkan analisis terhadap berbagai komplikasi yang dialami oleh pasien. Jumlah penderita penyakit kardiovaskular (CVD), termasuk jantung koroner, terus meningkat di seluruh dunia dan telah menjadi penyebab utama kematian secara global. Peningkatan ini dipengaruhi oleh berbagai faktor, seperti stres, hipertensi, kadar gula darah tinggi, serta faktor risiko lainnya[1]. Penyakit jantung merupakan salah satu penyebab utama tingginya angka kematian di Indonesia. Kondisi ini menjadi perhatian serius karena dapat menyerang siapa saja, terutama individu dengan gaya hidup kurang sehat atau yang memiliki riwayat penyakit tertentu. Gagal jantung merupakan penyebab utama kematian global, dengan sekitar 17,9 juta kematian per tahun menurut World Health Organization. Di Indonesia, data Riskesdas menunjukkan prevalensi penyakit jantung meningkat dari 0,5% (2013) menjadi 1,5% (2018), seiring pertambahan usia, pertumbuhan populasi, dan perubahan pola hidup masyarakat. Penelitian tahun 2021 menggunakan *Logistic Regression* pada 303 data *Cleveland Heart Disease* dan menunjukkan performa cukup baik (*sensitivity* 88,54% dan *specificity* 87,50%)[13]. Berbagai faktor memengaruhi munculnya gangguan jantung, salah satunya adalah pola hidup yang tidak baik, seperti kebiasaan makan yang kurang sehat, kurangnya aktivitas fisik, dan kebiasaan merokok. Faktor-faktor tersebut terbukti meningkatkan risiko penyakit jantung dan dapat menyebabkan komplikasi yang lebih berat[21]. Namun, jumlah data yang terbatas menjadi kendala dalam generalisasi model. Kondisi ini menekankan pentingnya penerapan metode diagnosis dini yang lebih efektif untuk mencegah terjadinya komplikasi yang lebih parah[22]. Penelitian ini bertujuan untuk mengimplementasikan metode seleksi fitur ANOVA F-Value guna mengidentifikasi atribut yang paling signifikan dalam prediksi penyakit jantung. Selanjutnya, dilakukan analisis komparatif terhadap dua algoritma machine learning, yaitu SVM dan *Random Forest*, untuk menentukan model yang paling efektif berdasarkan fitur yang telah terseleksi. Hasil evaluasi performa kedua algoritma ditampilkan melalui aplikasi berbasis Streamlit, sehingga proses analisis dan perbandingan akurasi dapat disajikan secara interaktif dan mudah dipahami.

II. METODE PENELITIAN

Desain penelitian dari setiap tahapan pada penelitian ini sebagai berikut: Gambar 1 menggambarkan tahapan penelitian.



Gambar 1. Tahapan Penelitian

2.1 Tahapan Pendahuluan

Tahapan ini merupakan langkah awal penelitian yang diawali dengan identifikasi masalah untuk memahami fenomena meningkatnya kasus penyakit jantung serta kurangnya model prediksi yang efisien. Batasan penelitian juga ditetapkan agar fokus pada penggunaan dataset “*Heart Disease Prediction*” dari Kaggle dengan pembahasan terbatas pada perbandingan performa kedua algoritma.

2.2 Studi Literatur

Setelah tahap pendahuluan, dilakukan tinjauan pustaka dengan mengumpulkan referensi dari jurnal, artikel, dan buku terkait penyakit jantung serta penerapan algoritma SVM dan Random Forest dalam prediksi berbasis machine learning. Tinjauan ini bertujuan mengidentifikasi kesenjangan penelitian sebelumnya dan menentukan metodologi yang tepat sebagai landasan pengembangan model prediksi.

2.3 Pengumpulan Data

Dataset *Heart Disease Prediction* dari Kaggle dipilih berdasarkan kelengkapan variabel, kualitas data, dan relevansinya. Dataset ini kemudian diunduh dalam format CSV dan dipersiapkan untuk proses analisis dan pelatihan model. Dataset *Heart Disease Prediction* berisi informasi medis dari berbagai individu yang digunakan untuk memprediksi risiko penyakit jantung. Dataset ini memuat data pasien dengan rentang usia dan jenis kelamin yang berbeda, serta mencakup variabel-variabel terkait kondisi kesehatan, gaya hidup, dan hasil pemeriksaan medis. Dataset terdiri dari 1.000 data pasien dan memiliki 16 atribut yang berupa data numerik dan kategorikal. Penjelasan rinci variabel dalam dataset ini disajikan dalam tabel 1. Variabel Independen dan tabel 2. Dependen.

Tabel 1. Variabel Independen

No	Variabel	Keterangan
1	Age	Usia individu dalam satuan tahun
2	Gender	Jenis kelamin (Laki-laki atau Perempuan)
3	Cholesterol	Tingkat kolesterol dalam mg/dL
4	Blood Pressure	Tekanan darah sistolik dalam mmHg
5	Heart Rate	Jumlah detak jantung per menit
6	Smoking	Status merokok (Tidak Pernah, Pernah, Sedang Merokok)
7	Alcohol Intake	Frekuensi konsumsi alkohol (Tidak Ada, Sedang, Berat)
8	Exercise Hours	Jumlah jam olahraga setiap minggu
9	Family History	Riwayat keluarga dengan penyakit jantung (Ya atau Tidak)
10	Diabetes	Status diabetes (Ya atau Tidak)
11	Obesity	Status obesitas (Ya atau Tidak)
12	Stress Level	Tingkat stres pada skala 1 sampai 10
13	Blood Sugar	Kadar gula darah puasa dalam mg/dL
14	Exercise include	Adanya angina akibat olahraga (Ya atau Tidak)
15	Chest Pain Type	Jenis nyeri dada (Angina Tipikal, Atipikal, Non-angina, atau Asimtomatik)

Tabel.2 Variabel Dependen

NO	Variable	Keterangan
16	Heart Disease	Variabel target yang menunjukkan keberadaan penyakit jantung (0: Tidak, 1: Ya)

2.4 Preprocessing

Tahap preprocessing meliputi pengecekan missing value, konversi variabel kategorikal ke numerik, serta normalisasi agar data berada pada skala yang sama. Selanjutnya dilakukan seleksi fitur menggunakan ANOVA F-Value (SelectKBest). Berdasarkan fitur terpilih, dibangun dua model menggunakan SVM dan Random Forest, lalu dievaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Model dengan performa terbaik kemudian diintegrasikan ke dalam aplikasi berbasis Streamlit. [11],[12],[13].

2.5 Streamlit

Pada tahap ini, Streamlit digunakan sebagai sarana untuk menampilkan hasil analisis dan prediksi yang telah diperoleh dari model. Model dengan performa terbaik, yaitu SVM atau Random Forest, diintegrasikan ke dalam Streamlit sehingga hasil prediksi dapat divisualisasikan secara interaktif. Melalui Streamlit, ditampilkan berbagai output seperti nilai akurasi, hasil klasifikasi, serta visualisasi data, dengan tampilan yang informatif dan mudah dipahami. Tahap ini tidak mencakup pembuatan aplikasi web, melainkan berfokus pada penyajian hasil menggunakan antarmuka Streamlit[14],[15].

2.6 Proses Evaluasi

Tahap ini merupakan proses evaluasi terhadap hasil analisis dan prediksi yang telah ditampilkan menggunakan Streamlit. [16],[17]. Pengujian dilakukan untuk memastikan bahwa model prediksi, seperti SVM atau Random Forest, memberikan hasil yang akurat dan konsisten. Selanjutnya, hasil pengujian dianalisis untuk menilai tingkat keakuratan dan efektivitas model dalam memprediksi risiko penyakit jantung. Berdasarkan analisis tersebut, disusun kesimpulan mengenai performa model serta rekomendasi untuk pengembangan penelitian di masa mendatang[18],[19],[20].

2.7 Pengolahan Data

Langkah sistematis yang dilakukan untuk mempersiapkan, mengolah, dan menganalisis data menggunakan algoritma SVM dan algoritma *Random Forest*.

2.7.1 Algoritma SVM

Adalah algoritma *supervised learning* yang digunakan untuk tugas klasifikasi dan regresi, baik pada data linear maupun non-linear. Prinsip kerjanya adalah menentukan hyperplane terbaik yang memisahkan antar kelas dengan memaksimalkan margin antara decision boundary dan titik data terdekat. SVM memiliki beberapa jenis, di antaranya Binary Class SVM, Multi Class SVM, dan Fuzzy SVM [7],[8],[9],[10].

2.7.2 Algoritma *Random Forest*

Adalah metode machine learning berbasis ensemble yang meningkatkan akurasi klasifikasi atau regresi dengan membangun sejumlah decision tree secara acak. Setiap pohon dibentuk dari node akar hingga node daun untuk menghasilkan prediksi, kemudian hasilnya digabungkan secara kolektif. Proses pembentukan pohon mirip metode CART, namun tanpa dilakukan pruning[4],[5],[6].

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} p_i^2 \quad (1)$$

$$Gini_{split} = \sum_{i=0}^{k-1} \binom{n_i}{n} Gini(S_i) \quad (2)$$

2.7.3 SelectBest Anova F value

ANOVA merupakan metode seleksi fitur pada teknik SelectKBest yang digunakan untuk menguji perbedaan rata-rata antar kelompok. Semakin tinggi nilai F, semakin besar pengaruh suatu fitur terhadap variabel target [2],[3].

$$F = \frac{\text{variance between group}}{\text{variance within group}} \quad (3)$$

$$\text{Variance between groups} = \frac{\sum_i^n n_i (\bar{Y}_{ij} - \bar{y})^2}{(k-1)} \quad (4)$$

$$\text{Variance within groups} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{ij} - \bar{y})^2}{(n-k)} \quad (5)$$

2.8 Confusion matrix

Adalah tabel yang digunakan untuk menilai kinerja model klasifikasi dengan cara membandingkan hasil prediksi model terhadap label sebenarnya dalam dataset[23]. Matriks ini memberikan informasi tentang jumlah prediksi yang benar maupun salah, serta jenis kesalahan yang terjadi dalam proses klasifikasi[24],[25].

Tabel 3. Confusion matrix

Prediksi Sesuai	TP	FP
Prediksi Tidak Sesuai	FN	TN

Metrik tabel 3. ini memberikan gambaran seberapa baik model dalam mengklasifikasikan data dengan benar serta seberapa efektif dalam menangani kesalahan prediksi.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \times 100\% \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

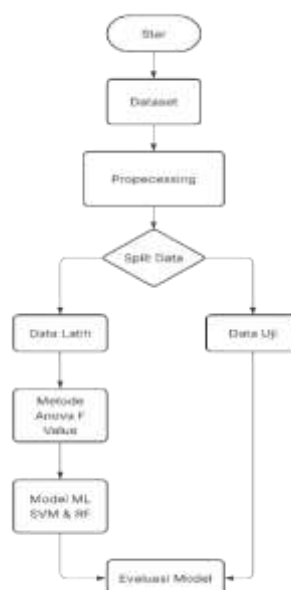
$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

III. HASIL DAN PEMBAHASAN

3.1 Perancangan Sistem

Proses implementasi pada algoritma *SVM* dan algoritma *Random Forest* menggunakan python untuk proses prediksi. Tahapan implementasi meliputi:

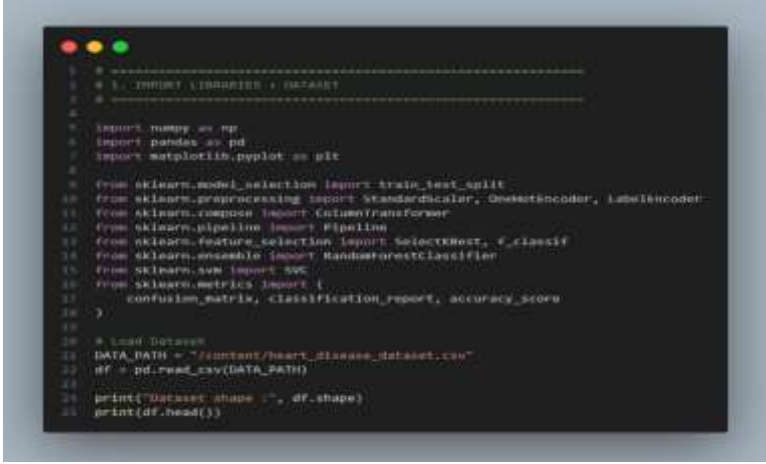
3.1.1 Rancang Model Machine Learning



Gambar 2. Rancang Model Machine Learning

Pada gambar 2. menggambarkan alur penelitian prediksi penyakit jantung yang dimulai dari pengumpulan dataset medis pasien, kemudian dilakukan preprocessing berupa pembersihan data, transformasi data kategorikal menjadi numerik, dan normalisasi

3.1.2 Import Dataset



```

1 # -----
2 # 1. IMPORT LIBRARIES + DATASET
3 # -----
4
5 import numpy as np
6 import pandas as pd
7 import matplotlib.pyplot as plt
8
9 from sklearn.model_selection import train_test_split
10 from sklearn.preprocessing import StandardScaler, OneHotEncoder, LabelEncoder
11 from sklearn.compose import ColumnTransformer
12 from sklearn.pipeline import Pipeline
13 from sklearn.feature_selection import SelectKBest, f_classif
14 from sklearn.ensemble import RandomForestClassifier
15 from sklearn.svm import SVC
16 from sklearn.metrics import (
17     confusion_matrix, classification_report, accuracy_score
18 )
19
20 # Load Dataset
21 DATA_PATH = "dataset/heart_disease_dataset.csv"
22 df = pd.read_csv(DATA_PATH)
23
24 print("Dataset shape :", df.shape)
25 print(df.head())

```

Gambar 3. Source code dataset

Gambar 3. menampilkan potongan kode program pada tahap import libraries dan pemanggilan dataset dalam penelitian prediksi penyakit jantung.

3.1.3 Preprocessing dan Split data

Tahap preprocessing data dilakukan serangkaian proses persiapan data sebelum digunakan pada model machine learning. Langkah ini diawali dengan menentukan kolom target, yaitu atribut terakhir pada dataset, kemudian memisahkan antara variabel input (X) dan variabel target (y).



```

1 # -----
2 # 2. PREPROCESSING
3 # -----
4
5 TARGET_COL = "heart Disease"
6
7 X = df.drop(columns=[TARGET_COL])
8 y = df[TARGET_COL]
9
10 # Encode target jika kategori
11 if y.dtype == 'O':
12     y = LabelEncoder().fit_transform(y)
13
14 # Pisahkan Numerik & Kategorikal
15 num_cols = X.select_dtypes(include=[np.number]).columns.tolist()
16 cat_cols = X.select_dtypes(include=['object', 'category']).columns.tolist()
17
18 # Isi Missing Value
19 for c in num_cols:
20     if X[c].isnull().any():
21         X[c].fillna(X[c].median(), inplace=True)
22
23 for c in cat_cols:
24     if X[c].isnull().any():
25         X[c].fillna(X[c].mode().iloc[0], inplace=True)
26
27 # Preprocessing Pipeline
28 numeric_t = Pipeline([("scaler", StandardScaler())])
29 categorical_t = Pipeline([("one", OneHotEncoder(handle_unknown="ignore"))])
30
31 preprocessor = ColumnTransformer([
32     ("num", numeric_t, num_cols),
33     ("cat", categorical_t, cat_cols)
34 ])
35
36 # Train-Test Split
37 X_train, X_test, y_train, y_test = train_test_split(
38     X, y, stratify=y, test_size=0.2, random_state=42
39 )
40
41 preprocessor.fit(X_train)
42 X_train_prep = preprocessor.transform(X_train)
43 X_test_prep = preprocessor.transform(X_test)
44
45 print("jumlah fitur setelah preprocessing:", X_train_prep.shape[1])
46

```

Gambar 4. Source code Preprocessing dan Split data

Berikut tabel 5. adalah 10 Variable yang dipilih menggunakan metode anova F value dengan F score tertinggi:

Tabel 5. Variabel Selection Form Anova F Value

Variable	F Score
Age	718.078783
Cholesterol	153.434140
Family History	0.916062
Gender	0.863290
Chest Pain Type	0.416951
Obesity	0.355781
Alcohol Intake	0.330061
Diabetes	0.273480
Exercise Hours	0.202008
Heart Rate	0.174159

3.1.5 Model Machine Learning SVM dan Random Forest



```

1 #
2 # 5. TRAINING MODEL (RANDOM FOREST & SVM)
3 #
4
5 # Random Forest
6 rf = RandomForestClassifier(
7     n_estimators=100,
8     max_depth=5,
9     min_samples_split=5,
10    min_samples_leaf=3,
11    random_state=42
12)
13
14 # SVM
15 svm = SVC(C=1, kernel='rbf', gamma='scale')
16
17 rf.fit(X_train_sel, y_train)
18 svm.fit(X_train_sel, y_train)
19
20 pred_rf = rf.predict(X_test_sel)
21 pred_svm = svm.predict(X_test_sel)
22
23 print("\n=== Random Forest Report ===")
24 print(classification_report(y_test, pred_rf))
25
26 print("\n=== SVM Report ===")
27 print(classification_report(y_test, pred_svm))

```

Gambar 6. Source Code Model Machine Learning SVM dan Random Forest

Pada tahap pelatihan model, pada gambar 6. Hasil evaluasi disajikan melalui *classification report* yang sehingga memudahkan dalam membandingkan kinerja kedua model.

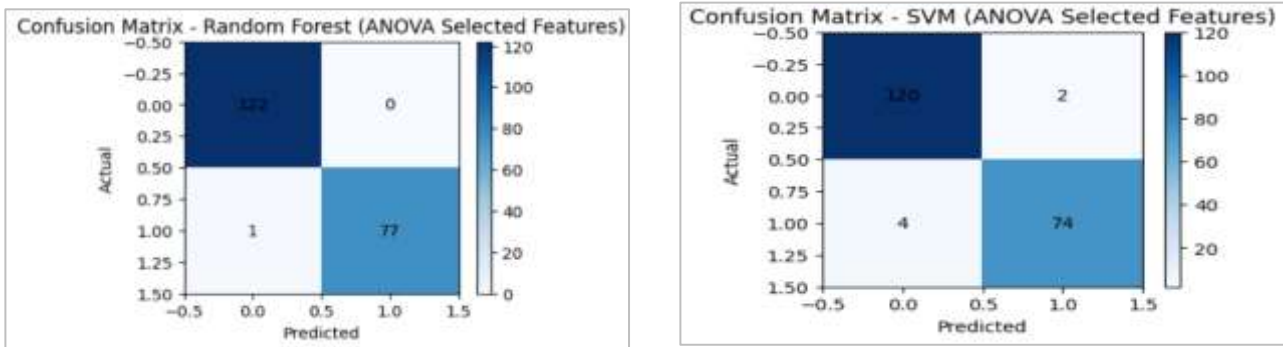
3.1.6 Hasil

=== RANDOM FOREST REPORT ===					=== SVM REPORT ===				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.99	1.00	1.00	122	0	0.97	0.98	0.98	122
1	1.00	0.99	0.99	78	1	0.97	0.95	0.96	78
accuracy			0.99	200	accuracy			0.97	200
macro avg	1.00	0.99	0.99	200	macro avg	0.97	0.97	0.97	200
weighted avg	1.00	0.99	0.99	200	weighted avg	0.97	0.97	0.97	200

Gambar 7. Hasil Akurasi Algoritma Random Forest dan SVM

Pada Gambar 7. dapat dilihat bahwa hasil *classification report* menunjukkan bahwa model *Random Forest* memiliki performa yang sangat tinggi dengan akurasi **99%**. *Random Forest* memberikan hasil prediksi yang sangat akurat, sedangkan model *SVM* menunjukkan performa yang sangat baik dengan akurasi 97%.

3.2 Implementasi Algoritma



Gambar 9. Hasil Confusion matrix model Algoritma Random Forest dan SVM

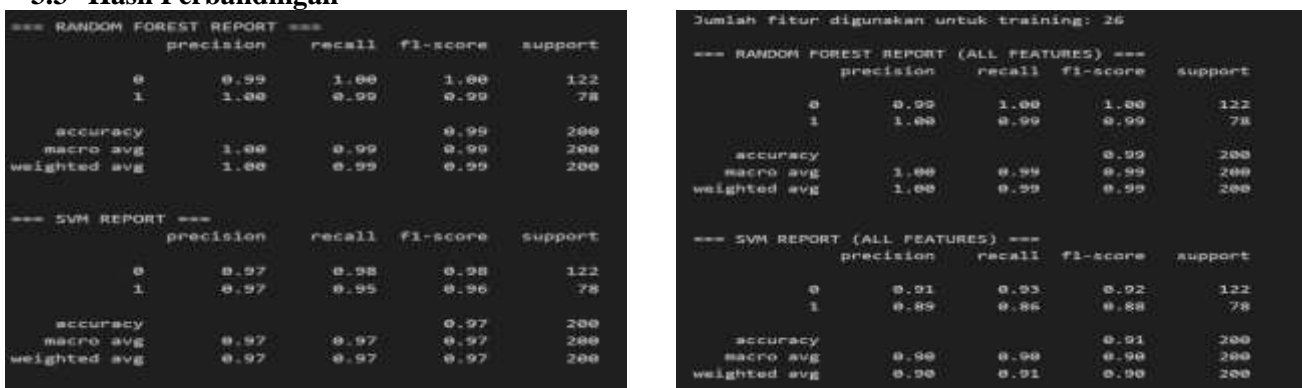
Gambar 9. Pada confusion matrix terlihat model algoritma **RF** bahwa kelas 0 berhasil diprediksi benar sebanyak 122 kali tanpa kesalahan, sedangkan kelas 1 diprediksi benar 77 kali dengan hanya satu kesalahan, sedang **SVM** berhasil memprediksi kelas 0 dengan benar sebanyak 120 data dan hanya melakukan 2 kesalahan, sedangkan untuk kelas 1 model memprediksi 74 data dengan benar dan juga mengalami 4 kesalahan.

Tabel 6. Perhitungan

Algoritma Random Forest	Algoritma SVM
$Accuracy = \frac{122 + 77}{122 + 77 + 1 + 1} \times 100\% = 99.15\%$	$Accuracy = \frac{74 + 120}{74 + 120 + 2 + 4} \times 100\% = 97\%$
$Precision = \frac{122}{122 + 0} = 0.9915$	$Precision = \frac{74}{74 + 2} = 0.9737$
$Recall = \frac{122}{122 + 1} = 0.9915$	$Recall = \frac{74}{74 + 4} = 0.9487$
$F1\ Score = 2 \times \frac{0.9915 \times 0.9915}{0.9915 + 0.9915} = 0.9703$	$F1\ Score = 2 \times \frac{0.9737 \times 0.9487}{0.9737 + 0.9487} = 0.9706$
$Accuracy = \frac{\sum_{i=1}^n I(\hat{y}_i = y_i)}{n} = \frac{199}{200} = 0.99$	$Accuracy = \frac{\sum_{i=1}^n I(\hat{y}_i = y_i)}{n} = \frac{194}{200} = 0.97$

Berdasarkan hasil perhitungan table 6. akurasi model adalah 0,99 atau 99%. Nilai ini diperoleh dari 199 prediksi benar dari total 200 data uji. Dengan tingkat kesalahan yang sangat kecil (hanya 1 data), model dapat dikatakan memiliki performa yang sangat baik dalam melakukan klasifikasi.

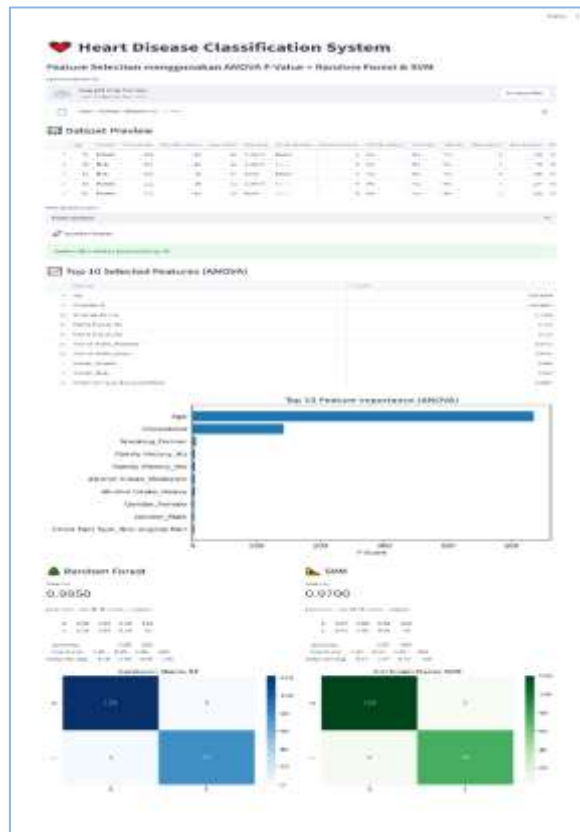
3.3 Hasil Perbandingan



Gambar 11. Perbandingan Akurasi *Random Forest* dan *SVM* dengan dan tanpa ANOVA *F-Value*

Gambar 11 menunjukkan bahwa dengan seleksi fitur ANOVA, Random Forest mencapai akurasi 0,99 dan SVM 0,97 dengan nilai precision, recall, dan F1-score yang tinggi dan seimbang. Namun tanpa ANOVA (menggunakan 26 fitur penuh), akurasi *Random Forest* tetap 0,99, sedangkan SVM turun menjadi 0,91 disertai penurunan metrik evaluasi, terutama pada kelas 1. Hal ini menunjukkan bahwa *Random Forest* lebih stabil terhadap banyak fitur, sementara SVM sangat terbantu oleh seleksi fitur karena sensitif terhadap fitur tidak relevan atau noise.

3.4 Tampilan Aplikasi



Gambar 13. Tampilan Streamlit

Gambar 13 menampilkan aplikasi *Heart Disease Classification System* berbasis Streamlit yang digunakan untuk klasifikasi penyakit jantung dengan metode seleksi fitur ANOVA F-Value serta algoritma Random Forest dan SVM. Setelah diproses, sistem menampilkan jumlah fitur hasil preprocessing serta 10 fitur terbaik berdasarkan nilai ANOVA F-Score dalam bentuk tabel dan grafik, yang menunjukkan bahwa Age dan Cholesterol merupakan fitur paling berpengaruh dalam prediksi penyakit jantung.

VI. KESIMPULAN

Seleksi fitur ANOVA F-Value efektif dalam mengidentifikasi atribut paling signifikan sehingga meningkatkan efisiensi dan akurasi model prediksi penyakit jantung. Kedua algoritma, SVM dan *Random Forest*, menunjukkan performa sangat baik, namun *Random Forest* lebih unggul dengan akurasi 99% dibandingkan 97% pada SVM serta metrik evaluasi yang lebih tinggi. Integrasi melalui Streamlit juga berhasil menampilkan hasil secara interaktif. Dengan demikian, kombinasi ANOVA F-Value dan *Random Forest* menjadi pendekatan paling optimal dalam penelitian ini.

REFERENSI

- [1]. Alfarizi, M. R., & Al-farish, M. Z. (2023). *Penggunaan python sebagai bahasa pemrograman untuk machine learning dan deep learning*. 2, 1–6.
- [2]. Adinulhaq, J. M., & Sam'an, M. (2023). Perbandingan Kinerja Akurasi Model Mesin Learning Untuk Prediksi Penyakit Jantung. *Jurnal Komputer Dan Teknologi Informasi*, 1(2), 48–55. <https://doi.org/10.26714/jkti.v1i2.12918>
- [3]. Akurasi, P., Forest, R., Anova, D., Dhani, A. A., Azhima, T., Siswa, Y., Pranoto, W. J., Pada, S., & Data, K. (2024). Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting Improving the Accuracy of Random Forest with ANOVA and SMOTE on Stunting Data Classification. 13(2), 264–272. <https://doi.org/10.34148/teknika.v13i2.875>
- [4]. Al Azhima, S. A. T., Darmawan, D., Arief Hakim, N. F., Kustiawan, I., Al Qibtiya, M., & Syafei, N. S. (2022). Hybrid Machine Learning Model untuk memprediksi Penyakit Jantung dengan Metode Logistic Regression dan Random Forest. In *Jurnal Teknologi Terpadu* (Vol. 8, Issue 1, pp. 40–46). <https://doi.org/10.54914/jtt.v8i1.539>
- [5]. Alhabib, I. (2022). Komparasi Metode Deep Learning, Naïve Bayes Dan Random Forest Untuk Prediksi Penyakit Jantung. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, 6(2), 176. <https://doi.org/10.51211/itbi.v6i2.1881> , no 5 jadi 1
- [6]. Amaliah, S., & Nusrang, M. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman

- Kopi Di Kedai Kopi Konijiwa Bantaeng. 4(2), 121–127. <https://doi.org/10.35580/variansiunm31>
- [7]. Anita, M., Yulianti, I. G. D., & Pasaribu, S. V. (2025). Klasifikasi Faktor Risiko Penyakit Jantung Menggunakan Machine Learning. In HOAQ (High Education of Organization Archive Quality): Jurnal Teknologi Informasi (Vol. 16, Issue 1, pp. 68–78). <https://doi.org/10.52972/hoaq.vol16no1.p68-78>
- [8]. Cnn-svm, M. (2022). Klasifikasi Jenis Kanker Kulit Menggunakan CNN-SVM. 2(2), 133–144.
- [9]. Dharma, F., & Wijaya, H. (2024). Analisis Data Penyakit Jantung Menggunakan Algoritma Decesion Tree dan Support Vector Machines. 19(2), 156–163.
- [10]. Handayani, F. (2021). Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. Jurnal Edukasi Dan Penelitian Informatika (JEPIN), 7(3), 329. <https://doi.org/10.26418/jp.v7i3.48053>
- [11]. Hasanah, Herliyani., & Nurmalitasari. (2023). Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung. Prosiding Seminar Nasional Teknologi Dan Sains, 2, 13–18.
- [12]. Hidayat, R., Sy, Y. S., Sujana, T., Husnah, M., & Saputra, H. T. (2024). Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine. 5(2), 161–168.
- [13]. Junifer Pangaribuan, J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. Journal Information System Development (ISD), 06(02), 1–10.
- [14]. Nasution, N., Hasan, M. A., & Bakri Nasution, F. (2025). Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset. In IT Journal Research and Development (Vol. 9, Issue 2, pp. 140–150). <https://doi.org/10.25299/itjrd.2025.17941>
- [15]. Nasution, N., Nasution, F., & Hasan, M. A. (2024). Heart Disease Risk Prediction: Evaluating Machine Learning Algorithms With Feature Reduction Using Lda. In JURTEKSI (Jurnal Teknologi dan Sistem Informasi) (Vol. 11, Issue 1, pp. 9–16). <https://doi.org/10.33330/jurteksi.v11i1.3498>
- [16]. Pal, M., & Parija, S. (2021). Prediction of Heart Diseases using Random Forest. Journal of Physics: Conference Series, 1817(1). <https://doi.org/10.1088/1742-6596/1817/1/012009>
- [17]. Purnomo, I. I. (2023). Perancangan Sistem Evaluasi Tingkat Kompetensi Alumni dengan Algoritma Support Vector Machine (SVM) Linear Method Berbasis Framework Streamlit. 8(3).
- [18]. Rani, M. C., Dewi, R. A., Azkia, F. D., Wahyudi, M., Sumanto, & Budiman, A. S. (2025). Perbandingan Algoritma Random Forest, Naive Bayes, Dan Neural Network Dalam Klasifikasi Penyakit Jantung. In Jurnal Sains Informatika Terapan (Vol. 4, Issue 2, pp. 77–84). <https://doi.org/10.62357/jsit.v4i2.609>
- [19]. Reddy, V. S. K., Meghana, P., Reddy, N. V. S., & Rao, B. A. (2022). Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers. Journal of Physics: Conference Series, 2161(1). <https://doi.org/10.1088/1742-6596/2161/1/012015>
- [20]. Rehman, M. U., Naseem, S., Butt, A. U. R., Mahmood, T., Khan, A. R., Khan, I., Khan, J., & Jung, Y. (2025). Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. Scientific Reports, 15(1), 1–15. <https://doi.org/10.1038/s41598-025-96437-1>
- [21]. Ridwan, R., Handayani, H. H., Lestari, S. A. P., & Cahyana, Y. (2025). Evaluasi Kinerja Algoritma Random Forest Dan Gradient Boosting Untuk Klasifikasi Penyakit Jantung. In Jurnal Komtika (Komputasi dan Informatika) (Vol. 9, Issue 1, pp. 112–124). <https://doi.org/10.31603/komtika.v9i1.13450>
- [22]. Riyanto, V., Destiana, H., Prihatin, T., & Wijaya, G. (2025). MENGOPTIMALKAN PREDIKSI GAGAL JANTUNG DENGAN KOMBINASI. 8(1), 103–111.
- [23]. Sahelvi, E., Cikita, P., Sapitri, R. M., Rahmaddeni, R., & Efrizoni, L. (2025). Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung. In MALCOM: Indonesian Journal of Machine Learning and Computer Science (Vol. 5, Issue 3, pp. 830–840). <https://doi.org/10.57152/malcom.v5i3.1972>
- [24]. Selayanti, N., Putri, S. A., Kristanaya, M., Azzahra, M. P., Navsih, M. G., & Hindrayani, K. M. (2024). Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung. In Prosiding Seminar Nasional Sains Data (Vol. 4, Issue 1, pp. 895–906). <https://doi.org/10.33005/senada.v4i1.376>
- [25]. Tangkelobo, E., Mayaut, W., Listanto, H., Binanto, I., & Sianipar, N. F. (2023). Perbandingan Algoritma Klasifikasi Random Forest , Gaussian Naive Bayes , dan K-Nearest untuk Data Tidak Seimbang dan Data yang diseimbangkan dengan metode Random Undersampling pada dataset LCMS Tanaman Keladi Tikus.