Vol. 6, No. 2, Agustus 2025, hal. 199-205

Jurnal Computer Science and Information Technology (CoSciTech)

p-ISSN: 2723-567X
e-ISSN: 2723-5661

http://ejurnal.umri.ac.id/index.php/coscitech/index

Perbandingan Model *Machine Learning* Untuk Klasifikasi Deteksi Penyakit Jantung

Tengku Fawwaz Fatihul Ihsan[1], Ilham Ramadhan[2], Davie Rizky Akbar[3], Edi Ismanto [4]
Email: [1]230401264@student.umri.ac.id, [2]230401265@student.umri.ac.id, [3]230401283@student.uac.id, [4]edi.ismanto@umri.ac.id

[1,2,3,4] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 25 Juli 2025 | Direvisi: - | Disetujui: 2 Agustus 2025

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, sehingga deteksi dini menjadi aspek penting dalam upaya pencegahan. Penelitian ini bertujuan membangun model prediksi risiko penyakit jantung berdasarkan data klinis pasien menggunakan algoritma *Random Forest*. Dataset yang digunakan terdiri dari 303 data dengan 13 fitur seperti tekanan darah, kolesterol, detak jantung maksimum, dan lain-lain, serta satu atribut target bertingkat. Proses pengolahan data mencakup pembersihan nilai tidak valid seperti tanda tanya ('?') yang diubah menjadi missing values, serta penghapusan data yang tidak lengkap untuk menjaga integritas dataset. Setelah melalui eksplorasi data dan analisis korelasi antar fitur, model dilatih menggunakan algoritma *Random Forest* karena kemampuannya dalam klasifikasi multikelas dan ketahanan terhadap overfitting. Hasil evaluasi awal menunjukkan model memiliki akurasi prediksi yang baik dengan skor mencapai 0,89. Penelitian ini membuktikan bahwa pendekatan machine learning berbasis *Random Forest* efektif dalam membantu proses identifikasi risiko penyakit jantung secara sistematis, sehingga berpotensi menjadi alat pendukung keputusan dalam bidang kesehatan preventif.

Kata kunci: *Penyakit jantung, Random forest, Machine learning, Gradient Boosting, Prediksi*

Comparison of Machine Learning Models for Classification and Detection of Heart Disease

Abstract

Heart disease is one of the leading causes of death in the world, so early detection is an important aspect in prevention efforts. This study aims to build a heart disease risk prediction model based on patient clinical data using the Random Forest algorithm. The dataset used consists of 303 data with 13 features such as blood pressure, cholesterol, maximum heart rate, and others, as well as one nested target attribute. The data processing process includes cleaning invalid values such as question marks ('?') which are changed to missing values, and deleting incomplete data to maintain the integrity of the dataset. After going through data exploration and correlation analysis between features, the model is trained using the Random Forest algorithm because of its ability in multiclass classification and resistance to overfitting. The initial evaluation results show that the model has good prediction accuracy with a score reaching 0.89. This study proves that the Random Forest-based machine learning approach is effective in helping the process of systematically identifying heart disease risks, so it has the potential to be a decision support tool in the field of preventive health.

Keywords: *Heart disease, Random forest, Machine learning, Gradient Boosting, Prediction*

1. PENDAHULUAN

Di masa ketatnya perkembangan AI dan *big data*. *Artificial Intellegence* merupakan salah satu hal yang sedang besar saat ini. *Machine Learning* yang merupakan salah satu cabang dari AI yang merupakan mesin mesin yang dikembangkan dan bisa bekerja

tanpa diatur oleh manusia. Proses *machine Learning*ini meliputi ilmu ilmu seperti statistika, matematika, dan analisis data. Dalam beberapa tahun terakhir, *machine Learning*(ML) telah diakui sebagai alat yang sangat ampuh untuk kemajuan teknologi. Meskipun gerakan yang menerapkan ML dan kecerdasan buatan (AI) untuk mengatasi masalah-masalah sosial dan global terus berkembang, masih diperlukan upaya bersama untuk mengidentifikasi cara terbaik menerapkan alat-alat ini untuk mengatasi perubahan iklim. Banyak praktisi ML ingin bertindak, tetapi tidak yakin bagaimana caranya. Di sisi lain, banyak bidang telah mulai secara aktif mencari masukan dari komunitas *machine learning*[1]. *Deep Learning* merupakan subbidang pembelajaran mesin, sedangkan pembelajaran mesin merupakan subbidang AI. Hasilnya, *Machine Learning* dan *Deep Learning* digunakan untuk menciptakan sistem deteksi intrusi yang efisien dan efektif. Makalah ini memberikan gambaran umum tentang aplikasi dan pendekatan pembelajaran mesin dan pembelajaran mendalam dalam sistem deteksi intrusi dengan berkonsentrasi pada teknologi, metodologi, dan implementasi keamanan jaringan[2].

Selama dekade terakhir, kedua istilah tersebut, kecerdasan buatan (AI) dan *machine Learning*(ML), telah menikmati peningkatan popularitas dalam penelitian sistem informasi (IS). Sebuah analisis jurnal "*AIS Senior Scholars' Basket*" sejak tahun 2000,2 menggambarkan bagaimana kemunculan kedua istilah tersebut meningkat dalam judul, abstrak, dan kata kunci[3].

Setiap tahun, *American Heart Association*, bekerja sama dengan *National Institutes of Health* dan lembaga pemerintah lainnya, mengumpulkan statistik terkini terkait penyakit jantung, stroke, dan faktor risiko kardiovaskular dalam Life's Essential 8 milik AHA dalam satu dokumen[4]. Penyakit jantung sering terjadi di sekitaran masyarakat, penyebab nya bisa muncul dari berbagai penyebab, seperti detak jantung yang berlebihan, kebiasaan tidur yang berbeda, ukuran stress, dan umur.

Random Forest adalah salah satu jenis metode bagging yang memiliki cara kerja dengan membangkitkan sejumlah pohon dari data sampel dimana pembuatan satu pohon selama proses pelatihan tidak bergantung pada pohon sebelumnya dan kemudian keputusan diambil berdasarkan yang paling banyak. suara. Dua konsep yang mendasari *Random Forest* adalah membangun komposit pohon melalui *bagging* dengan mengganti dan memilih fitur secara acak untuk setiap pohon yang dibangun[5]. *Random Forest* merupakan metode pembelajaran *ensemble* dengan menggabungkan beberapa pohon keputusan sehingga didapatkan hasil akurasi yang akurat.Beberapa *Random Forest* yang dilaporkan dalam literatur secara konsisten memiliki galat generalisasi yang lebih rendah daripada yang lain. Misalnya, pemilihan split acak (Dieterrich, 1998) lebih baik daripada *bagging*. Pengenalan derau acak ke dalam keluaran oleh *Breiman* (Breiman, 1998c) juga lebih baik. Namun, tidak satu pun dari ketiga hutan ini yang bekerja sebaik *Adaboost* (Freund & Schapire, 1996) atau algoritme lain yang bekerja dengan pembobotan ulang adaptif (arcing) dari set pelatihan (lihat Breiman, 1998b; Dieterrich, 1998; Bauer & Kohavi, 1999). mempertahankan kekuatan[6].

Banyak eksperimen dan simulasi menghasilkan data dalam jumlah besar. Pemanfaatan data ini untuk penelitian fyang lebih baik tentang pembakaran telah menjadi tantangan dan peluang penelitian baru. Untungnya, *machine Learning*(ML) menyediakan teknik berbasis data canggih untuk mengekstrak informasi dari data besar dan membantu mengungkap mekanisme pembakaran yang mendasarinya[7].

Seiring dengan menjamurnya model *machine Learning* dalam berbagai aspek kehidupan kita, muncul kekhawatiran mengenai kemampuannya untuk menimbulkan bahaya. Dalam bidang kedokteran, kegembiraan mengenai kinerja *machine Learning* tingkat manusia untuk kesehatan diimbangi dengan masalah etika, seperti potensi alat-alat ini untuk memperburuk kesenjangan kesehatan yang ada. Misalnya, penelitian terkini telah menunjukkan bahwa model prediksi klinis mutakhir berkinerja buruk pada wanita, kelompok minoritas etnis dan ras, dan mereka yang memiliki asuransi publik. Sementara dalam pembelajaran *machine learning*, pakar harus menentukan representasi yang diperlukan, dalam pembelajaran mendalam, representasi diidentifikasi secara otomatis melalui penggunaan algoritma pembelajaran mendalam[2]. Semakin banyak literatur yang membahas implikasi sosial dari *machine Learning* dan teknologi. Beberapa dari karya ini, yang disebut studi data kritis, berasal dari perspektif ilmu sosial , sedangkan karya lainnya berasal dari perspektif teknis dan ilmu komputer . Meskipun ada kajian yang membahas implikasi sosial dan keadilan algoritmik secara umum, hanya sedikit penelitian yang membahas tentang kesehatan, *machine learning*, dan keadilan , meskipun ada potensi dampak hidup-mati dari model machine learning[8].

Dalam percobaan dengan *Random Forest*, *bagging* digunakan bersamaan dengan pemilihan fitur acak. Setiap set pelatihan baru diambil, dengan penggantian, dari set pelatihan asli. Kemudian pohon ditanam pada set pelatihan baru menggunakan pemilihan fitur acak. Pohon yang tumbuh tidak dipangkas. *Random Forest* paling sederhana dengan fitur acak dibentuk dengan memilih secara acak, di setiap simpul, sekelompok kecil variabel input untuk dibagi. Dalam implementasi hutan acak, perhitungan untuk setiap variabel kategoris hanya melibatkan pemilihan subset kategori secara acak[6]. *Random Forest* juga memiliki kemampuan antara lain adalah bisa menangani data berdimensi tinggi dan mengurangi *overfitting*[9]. *Random Forest* juga dapat menghasilkan error yang rendah, menghasilkan klasifikasi yang baik dan dapat mengatasi data *training* yang besar[10].

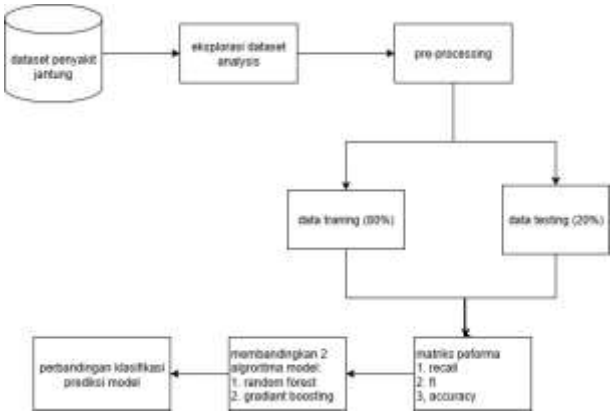
Dalam Gradient Boosting, metode pembelajaran ini bekerja dengan menciptakan model dasar baru secara berurutan yang mempelajari bagian konsep yang belum 'dipahami' oleh pembelajar sebelumnya. Permasalahan ini diajukan sebagai permasalahan optimasi di mana keluaran setiap model baru dibangun agar berkorelasi dengan gradien negatif fungsi kerugian iterasi sebelumnya[11]. Perbedaan utama antara peningkatan dan peningkatan gradien adalah bagaimana kedua algoritma memperbarui model (pembelajar lemah) dari prediksi yang salah[12]. Algoritma *Gradient Boosting* (GB) juga menciptakan beberapa pembelajar lemah untuk membentuk satu pembelajar kuat. Sebuah pembelajar lemah pertama-tama dibangun untuk memprediksi variabel keluaran. Kemudian, pada iterasi berikutnya, pembelajar lemah lainnya dibangun, namun kali ini pembelajar lemah tersebut dibangun untuk mempelajari kesalahan residual dari pembelajar lemah sebelumnya[13]. Di antara pemodelan pohon keputusan, *Gradient Boosting Machine* (GBM) telah mengalami lonjakan popularitas yang kuat dalam

beberapa tahun terakhir, didorong oleh hasil yang sangat baik dalam kompetisi ilmu data dan kinerja mutakhir dalam pemodelan data tabular [14]. Peningkatan gradien menyesuaikan bobot dengan penggunaan gradien (arah dalam fungsi kerugian) menggunakan algoritma yang disebut Penurunan gradien, yang secara berulang mengoptimalkan kerugian model dengan memperbarui bobot[12]. Reputasi *Gradient boosting* membuat para sains data menggunakan model tersebut untuk mengolah data tipe klasifikasi.

Masalah umum dengan peningkatan gradien adalah pemilihan hiperparameter seperti jumlah pohon dan parameter yang mengatur struktur pohon. Kami mengusulkan pemilihan hiperparameter adaptif dengan validasi silang berbasis deviasi. Hal ini tidak mudah karena kuantitas yang diinginkan, yaitu kuantil kondisional ekstrem, tidak memiliki hubungan yang jelas dengan deviasi[15].

2. METODE PENELITIAN

Metode yang akan digunakan adalah model *machine Learning* yang mengacu pada dua dari berbagai model, yaitu model *Random Forest* dan *Gradient Boosting*. Masalah pada penelitian ialah bagaimana cara kita untuk memprediksi berdasarkan klasifikasi dari masing masing fitur atau kolom pada sebuah dataset, yang dimana dataset itu adalah dataset tentang penyakit jantung yang memiliki banyak kegunaan jika kita mengolah data dengan baik. Berikut adalah diagram bagaimana cara memproses dataset penyakit jantung pada penelitian ini.



Gambar 1. Diagram Alur Penelitian

2.1. Dataset

Dataset dalam penelitian yang akan digunakan adalah dataset penyakit jantung. Permasalahan pada penelitian ini adalah pemodelan *Machine Learning* jenis apakah yang sangat akurat dalam prediksi berdasarkan klasifikasi. Se efektif apapun caranya, data ini bisa memberikan manfaat dalam dunia medis didalam beberapa aspek, seperti aspek teknologi.Dengan penggunaan dan pengolahan data yang bisa menjadi lebih efektif dibandingkan teknik manual. Peran data disini adalah dapat membantu sebisa mungkin pada departemen kesehatan, walau teknik konvensional lebih baik dan akurat. Dataset yang akan di teliti ini memiliki 300 lebih dari data pasien yang memiliki riwayat penyakit jantung. Dengan fitur atau kolom yang ada didalam data ini, masalah untuk membuat pemodelan prediksi klasifikasi akan berjalan dengan baik.

2.2. *EDA (Exploratory Data Analysis)* dan *Pre-Processing*

Dengan melakukan *EDA* atau *Exploratory Data Analysis*, kami akan memahami isi dari data dengan melihat atau memanggil 5 sampel awal atau acak dari data dengan google collab, berikut adalah sintaks nya.

Kode Program

```
from google.collab import files
uploaded = files.upload()
df = pd.read_csv('Heart_Disease_Data.csv')
print(df.head())
```

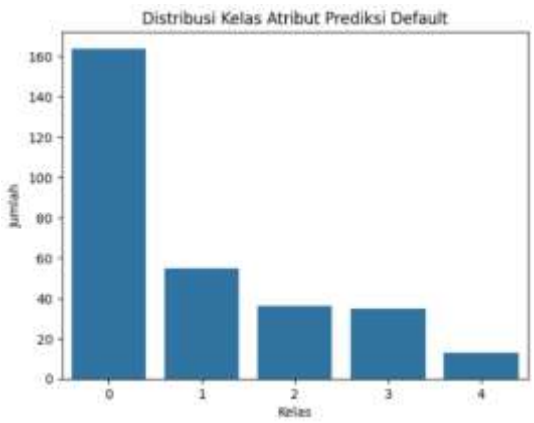
Sintaks diatas akan mengimpor dataset yang kita tambahkan kepada *google collab*, kemudian memanggil variabel df yang akan membaca data tersebut. Kemudian menampilkan output dari variabel dataset dengan sintaks head yang akan menampilkan 5 data awal sampel untuk di eksplorasi dan dipahami.

Setelah mengeksplor data, kami memahami bawah segala fitur atau kolom yang disajikan didalam data memiliki fungsi masing masing yang akan memberikan kerumitan pengklasifikasi-an kepada hasil dari target. Kami juga menyadari bahwa target atau tujuan dari dataset ini memiliki skala prediksi yang sangat bervariasi. Dengan target yang sangat bervariasi, juga dengan data yang tidak terlalu banyak, maka kami akan mengubah target menjadi *multi class*, supaya hasil dari prediksi berdasarkan jumlah data yang akan dilatih bisa memiliki akurasi yang lebih akurat.

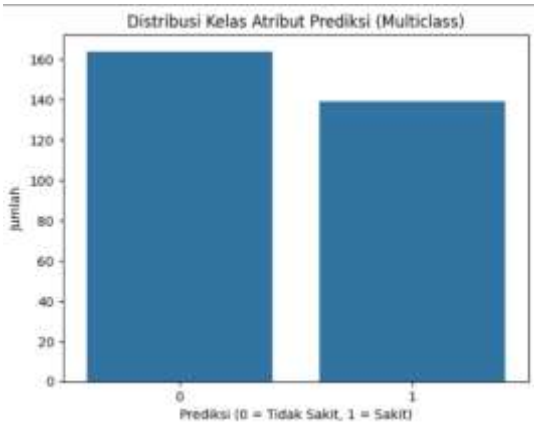
Kode Program

```
df['pred_attribute'] = df['pred_attribute'].apply(lambda x: 1 if x > 0 else 0)
```

Dengan membuat tipe target menjadi multi class, yaitu dengan membat skala yang lebih dari 0 akan jadi 1. Berikut adalah perbandingan skala prediksi dari keseluruhan isi data dari sebelum mengaplikasikan multiclass pada target data.



Gambar 2. Tampilan Target Sebelum Multiclass



Gambar 3. Tampilan Target Setelah Multiclass

Dengan mengpalikasikan multiclass dan melihat fungsi fungsi tiap fitur data sudah dipahami. Maka metode pemrosesan data akan lanjut ke pembersihan data. Pembersihan data dilakuka dengan mengidentifikasi tiap tipe data dari fitur atau kolom didalam data, dengan kode program berikut.

Kode Program

```
df.info()
```

Sintaks atau kode program diatas akan mengidentifikasi tipe data dari semua fitur pada data. Fitur ‘ca’ dan ‘thal’ teridentifikasi sebagai objek, sedangkan fitur lain memiliki tipe data integer, kita akan menyesuaikannya. Ini adalah langkah langkah encoding yaitu mengubah bentuk dari dataset untuk siap latih model machine leaning. Sebelum mengubahnya menjadi numerik tipe data, kita akan mengecek nilai unik atau nilai berbeda dari kedua kolom ini.

Kode Program

```
print(df['ca'].unique())
print(df['thal'].unique())
```

Hasil dari mencari nilai unik menampilkan bahwa ada nilai ‘nan’ yang bisa diidentifikasikan sebagai data kosong atau tidak ada isi pada data. Kami akan menambahkan imputasi yang akan mengisi data kosong pada kedua fitur dengan nilai yang paling sering muncul atau modus dari masing masing kolom.

Kode Program

```
for col in ['ca', 'thal']:
    if df[col].isnull().any():
        mode_val = df[col].mode()[0]
        df[col].fillna(mode_val, inplace=True)
        print(f"Nilai hilang di '{col}' diisi dengan modus: {mode_val}")
```

Disini kami juga melakukan normalisasi untuk seluruh isi dataset. Normalisasi digunakan jika dataset yang kurang seimbang atau memiliki sedikit data. Beberapa nilai nilai yang terhitung besar pada beberapa fitur seperti kadar kolesterol yang tergolong memiliki skala nilai unik yang lebih besar dibanding umur. Karena itulah normalisasi pada dataset ini diperlukan.

Setelah melakukan *Explorary Data Analysis* seperti eksplorasi data dan pembersihan data, Maka data siap untuk dilatih dengan kedua model yang akan dibandingkan. Tahap selanjutnya adalah pembagian dataset yang akan diproses kepada dua model *Machine Learning*.

2.3. Pembagian Dataset

Pembagian dataset akan dilakukan setelah melakukan *pre-processing*. Dataset akan dibagi menjadi dua, yaitu data latih dan data uji. Hal ini dilakukan untuk membagi isi dataset yang bisa digunakan untuk melatih model, dan juga untuk mengevaluasi hasil peforma dari model *Machine Learning* dengan data uji.

Kode Program

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

Sintaks diatas akan membagi skala data yang akan diproses atau dilatih dengan model. Dengan mengatur data *test* atau *test_size* yaitu 20 % atau 0,2, dan *random state* yang akan mengambil data sesuai urutan *random*. Dengan menagambil *library sklearn model* yang akan membaca sintaks untuk pembagian data. Data akan siap untuk dilatih dengan model machine learning.

2.4. Pelatihan Data dengan Model

Dataset akan dikembangkan atau dilatih dengan dua model tipe klasifikasi berbasis dari pengolahan pohon atau *tree-based* yang dimana fokus model akan memproses dengan *decision tree*. Dua model yang akan digunakan adalah *Random Forest* dan *Gradient Boosting.* Kedua model akan mengolah dataset yang sudah di proses dan siap uji latih model. Pelatihan kedua model juga akan didukung oleh *GridSearchCV*. *GridSearchCV* merupakan teknik pencarian otomatis yang bisa membantu pelatihan data untuk mencari kombinasi berdasarkan hyperparameter terbaik dari model model machine learning.

Random Forest adalah algoritma model *machine Learning* yang secara proses akan membuat banyak decision secara acak tetapi sesuai dengan data sampel, kemudian akan menggabungkan semua hasil decision tree tersebut untuk menghasilkan prediksi dan aktual yang lebih stabil dan akurat. Pada kasus data klasifikasi, model *Random Forest* akan mengambil hasil mayoritas dari semua *decision tree* yang dibuat.

Gradient Boosting adalah model algoritma *Machine Learning*yang bisa mengolah data tipe klasifikasi dengan cara membuat *decision tree* secara bertahap dari sampel data yang diambil, dengan memperbaiki kesalahan atau kekurangan dari *decision tree* sebelumnya secara lanjutan. Disebut sebagai *boosting* karena cara pelatihan yang akan meningkatkan peforma latihan secara bertahap. Kedua model ini akan dilatih dengan dataset yang sudah siap untuk diuji.

2.5. Output Hasil Pelatihan

Evaluasi atau penghitungan hasil akhir dari kedua model klasifikasi dilakukan dengan cara mengukur presisi, *recall*, dan *f1-score*. Dengan presisi yang menghitung akurat prediksi dari target data, kemudian *recall* atau sensifitas sebuah model dalam menangkap kasus positif, dan *f1-score* yaitu rata rata dari keseluruhan ukuran hasil akurasi. Dan juga masing masing model akan di nilai berdasarkan *macro average* yang menghitung rata rata semua hasil sampel data berdasarkan jumlah kelas atau fitur, dan berdasarkan *weighted average* yaitu rata rata dari semua fitur berdasarkan nilai support atau seluruh data yang di rata-rata kan dengan jumlah data.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari penelitian yang telah dilakukan untuk memprediksi klasifikasi penyakit jantung dari data pasien. Peniliti menggunakan dua model algoritma diantaranya *Random Forest* dan *Gradient Boosting*. Evaluasi performa dari model menggunakan metrik *precision*, *recall*, *fl-score,* dan *support* untuk mengukur kedua model tersebut dalam menganalisis data pasien tersebut.

Table 1. Evaluasi Model *Random Forest*

No	Precision	Recall	Fl-score	Support
0	0.93	0.85	0.89	33
1	0.84	0.93	0.88	28
Accuracy			0.89	61

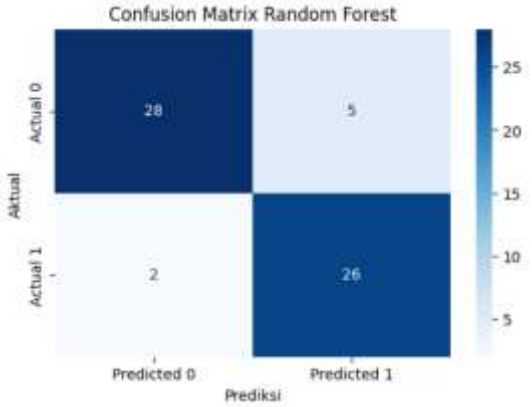
Macro avg	0.89	0.89	0.89	61
Weighted avg	0.89	0.89	0.89	61

Berdasarkan dari Table 1 model *Random Forest* didapat bahwa nilai *precision* untuk yang tidak memiliki penyakit jantung (0) adalah 0,93, yang memiliki penyakit jantung (1) 0.84, dan 0.89 untuk *macro avg* dan *weigted avg.* Nilai *recall* untuk yang tidak memiliki penyakit jantung (0) adalah 0,85, yang memiliki penyakit jantung (1) 0.93, dan 0.89 untuk *macro avg* dan *weigted avg.* Nilai *Fl-score* untuk yang tidak memiliki penyakit jantung (0) adalah 0,89, yang memiliki penyakit jantung (1) 0.89, dan 0.89 untuk *accuracy, macro avg, dan weigted avg.* Nilai *Support* untuk yang tidak memiliki penyakit jantung (0) adalah 33, yang memiliki penyakit jantung (1) 28, dan 61 untuk *accuracy*, *macro avg*, dan *weigted avg.*

Table 2. Evaluasi Model *Gradient Boosting*

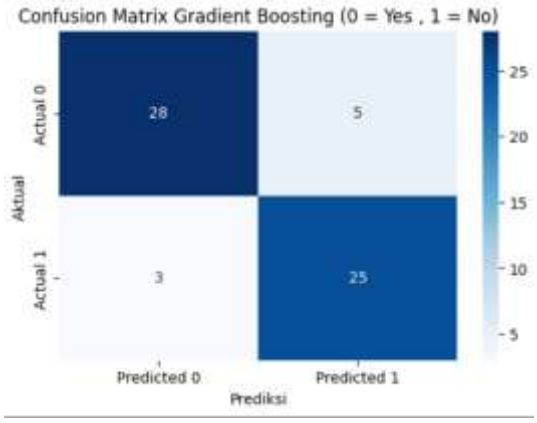
No	Precision	Recall	Fl-score	Support
0	0.90	0.85	0.88	33
1	0.83	0.89	0.86	28
Accuracy			0.87	61
Macro avg	0.87	0.87	0.87	61
Weighted avg	0.87	0.87	0.87	61

Untuk Table 2 model *Gradient Boosting* didapat bahawa nilai *precision* untuk yang tidak memiliki penyakit jantung (0) adalah 0,90, yang memiliki penyakit jantung (1) 0.83, dan 0.87 untuk *macro avg dan weigted avg.* Nilai *recall* untuk yang tidak memiliki penyakit jantung (0) adalah 0,85, yang memiliki penyakit jantung (1) 0.89, dan 0.87 untuk *macro avg dan weigted avg.* Nilai *Fl-score* untuk yang tidak memiliki penyakit jantung (0) adalah 0,88, yang memiliki penyakit jantung (1) 0.86, dan 0.87 untuk *accuracy, macro avg, dan weigted avg.* Nilai *Support* untuk yang tidak memiliki penyakit jantung (0) adalah 33, yang memiliki penyakit jantung (1) 28, dan 61 untuk *accuracy, macro avg, dan weigted avg.*



Gambar 4. *Matrix Confusion Random Forest*

Dari Gambar 4 *matrix confusion Random Forest* didapat actual 0 predicted 0 yang berarti 28 kasus yang kelasnya 0 (tidak terkena penyakit) dan model juga benar memprediksinya sebagai kelas 0 maka kasus tersebut dianggap benar (*true*). Actual 1 predicted 1 yang berarti 26 kasus yang kelasnya 1 (terkena penyakit) model juga benar memprediksinya sebagai kelas 1 maka kasus tersebut dianggap benar.



Gambar 5. Matrix Confusion Gradient Boosting

Didapat dari dari Gambar 5 *matrix confusion Random Forest* didapat actual 0 predicted 0 yang berarti 28 kasus yang kelasnya 0 (tidak terkena penyakit) dan model juga benar memprediksinya sebagai kelas 0 maka kasus tersebut dianggap benar (*true*). *Actual* 1 *predicted* 1 yang berarti 25 kasus yang kelasnya 1 (terkena penyakit) model juga benar memprediksinya sebagai kelas 1 maka kasus tersebut dianggap benar.

4. KESIMPULAN

Dari hasil penelitian permasalahan penyakit jantung dan eksperimen menggunakan dua model yaitu *Random Forest* dan *Gradient Boosting* didapat hasilnya bahwa model *Random Forest* mendapatkan akurasi yang lebih tinggi dari pada model *Gradient Boosting.* Dengan evaluasi nilai *precision, recall, fl-score* yang lebih tinggi. Pada *matrix confusion* yang memiliki hasil *actual true* dan *actual false* yang lebih banyak dari pada *Gradient Boosting* dengan true 28 kasus untuk yang tidak terkena penyakit dan 26 kasus untuk yang terkena penyakit. Diharapkan dengan pemodelan yang lebih akurat dan baik akan memberikan manfaat dari dataset pada permasalahan ini.

DAFTAR PUSTAKA

[1] D. Rolnick *et al.*, "Tackling Climate Change with Machine Learning," *ACM Comput. Surv.*, vol. 55, no. 2, 2023, doi: 10.1145/3485128.
[2] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine Learning and Deep Learning Approaches for CyberSecurity: A Review," *IEEE Access*, vol. 10, no. Ml, pp. 19572–19585, 2022, doi: 10.1109/ACCESS.2022.3151248.
[3] N. Kühl, M. Schemmer, M. Goutier, and G. Satzger, "Artificial intelligence and machine learning," *Electron. Mark.*, vol. 32, no. 4, pp. 2235–2244, 2022, doi: 10.1007/s12525-022-00598-0.
[4] C. W. Tsao *et al.*, *Heart Disease and Stroke Statistics - 2023 Update: A Report from the American Heart Association*, vol. 147, no. 8. 2023. doi: 10.1161/CIR.0000000000001123.
[5] M. A. Abubakar, M. Muliadi, A. Farmadi, R. Herteno, and R. Ramadhani, "Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Inform.*, vol. 10, no. 1, pp. 13–18, 2023, doi: 10.31294/inf.v10i1.14531.
[6] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0_35.
[7] L. Zhou, Y. Song, W. Ji, and H. Wei, "Machine learning for combustion," *Energy AI*, vol. 7, p. 100128, 2022, doi: 10.1016/j.egyai.2021.100128.
[8] S. Kushwaha, R. Srivastava, H. Vats, and P. Khanna, "Machine learning in healthcare," *Mach. Learn. Soc. Improv. Mod. Prog.*, pp. 50–70, 2022, doi: 10.4018/978-1-6684-4045-2.ch003.
[9] Fitri Handayani and Reny Medikawati Taufiq, "Komparasi Algoritma Menggunakan Teknik Smote Dalam Melakukan Klasifikasi Penyakit Stroke Otak," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 5, no. 2, pp. 367–372, 2024, doi: 10.37859/coscitech.v5i2.7439.
[10] J. Al Amien, Yoze Rizki, and Mukhlis Ali Rahman Nasution, "Implementasi Adasyn Untuk Imbalance Data Pada Dataset UNSW-NB15 Adasyn Implementation For Data Imbalance on UNSW-NB15 Dataset," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 242–248, 2022, doi: 10.37859/coscitech.v3i3.4339.
[11] S. Emami and G. Martínez-Muñoz, "Condensed-gradient boosting," *Int. J. Mach. Learn. Cybern.*, vol. 16, no. 1, pp. 687–701, 2025, doi: 10.1007/s13042-024-02279-0.
[12] S. Malik, R. Harode, and A. Singh Kunwar, "XGBoost: a deep dive into boosting," *Simon Fraser Univ.*, no. February, pp. 1–21, 2020, doi: 10.13140/RG.2.2.15243.64803.
[13] L. W. Rizkallah, "Enhancing the performance of gradient boosting trees on regression problems," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01071-3.
[14] D. Boldini, F. Grisoni, D. Kuhn, L. Friedrich, and S. A. Sieber, "Practical guidelines for the use of gradient boosting for molecular property prediction," *J. Cheminform.*, vol. 15, no. 1, pp. 1–13, 2023, doi: 10.1186/s13321-023-00743-7.
[15] J. Velthoen, C. Dombry, J. J. Cai, and S. Engelke, "Gradient boosting for extreme quantile regression," *Extremes*, vol. 26, no. 4, pp. 639–667, 2023, doi: 10.1007/s10687-023-00473-x.