

Optimasi Prediksi Kematian pada Gagal Jantung: Analisis Perbandingan Algoritma Pembelajaran Ensemble dan Teknik Penyeimbangan Data pada Dataset

Andhar Siraj Munir^{a1}, Retno Waluyo^{a2}

^aFakultas Ilmu Komputer, Universitas Amikom Purwokerto

Jl. Letjend Pol. Soemarto No.127, Watumas, Purwanegara, Kec. Purwokerto Utara, Kabupaten Banyumas, Jawa

Tengah 53127

¹andhar.moenir@gmail.com

²waluyo@amikompurwokerto.ac.id

Abstrak

Penyakit jantung merupakan penyebab utama kematian di seluruh dunia, termasuk di Indonesia. Identifikasi penyakit kardiovaskular (CVD) memerlukan pertimbangan berbagai faktor, seperti tekanan darah tinggi, kadar kolesterol, diabetes, dan lainnya, dengan gejala yang dapat bervariasi antar jenis kelamin. Meskipun angiografi dianggap metode yang akurat, biayanya tinggi dan kurang terjangkau oleh keluarga berpendapatan rendah. Biaya penyakit kardiovaskular juga memberikan dampak finansial signifikan pada sistem kesehatan. Dalam upaya untuk meningkatkan prediksi penyakit jantung, penelitian ini menggunakan metode ensemble learning, seperti Random Forest, Bagging, Adaboost, Gradient Boosting, dan XGBoost, dengan penyetelan hyperparameter. Eksperimen pada dataset gagal jantung menunjukkan bahwa penerapan teknik Synthetic Minority Over-sampling Technique (SMOTE) pada algoritma Extreme Gradient Boosting (XGB) memberikan hasil terbaik, mencapai akurasi 88.9%, F1-score 87.7%, dan Matthews Correlation Coefficient (MCC) 75.8%. Penggunaan metode balancing data, seperti SMOTE, ROS, dan RUS, secara signifikan memengaruhi performa algoritma, menyoroti pentingnya pemilihan metode sesuai dengan karakteristik dataset. Hasil ini memiliki implikasi penting dalam meningkatkan prediksi dan manajemen risiko kematian pada pasien gagal jantung secara dini dan lebih hemat biaya.

Kata kunci: Gagal Jantung, Prediksi Kematian, Algoritma Pembelajaran Ensemble, Teknik Penyeimbangan Data

Optimizing Heart Failure Mortality Prediction: Comparative Analysis of Ensemble Learning Algorithms and Data Balancing Techniques on a Dataset

Abstract

Heart disease is a leading cause of death worldwide, including in Indonesia. Identifying cardiovascular disease (CVD) involves considering various factors such as high blood pressure, cholesterol levels, diabetes, and others, with symptoms that can vary between genders. Although angiography is considered an accurate method, its cost is high and less accessible to low-income families. The financial impact of cardiovascular disease is also significant on the healthcare system. In an effort to improve heart disease prediction, this research employs ensemble learning methods such as Random Forest, Bagging, Adaboost, Gradient Boosting, and XGBoost, with hyperparameter tuning. Experiments on heart failure datasets show that applying the Synthetic Minority Over-sampling Technique (SMOTE) to the Extreme Gradient Boosting (XGB) algorithm yields the best results, achieving an accuracy of 88.9%, F1-score of 87.7%, and Matthews Correlation Coefficient (MCC) of 75.8%. The use of data balancing methods, such as SMOTE, ROS, and RUS, significantly impacts algorithm performance, highlighting the importance of method selection based on dataset characteristics. These findings have important implications for improving early prediction and management of death risk in heart failure patients, in a cost-effective manner.

Keywords: Heart Failure, Mortality Prediction, Ensemble Learning Algorithms, Data Balancing Techniques

I. PENDAHULUAN

Berdasarkan data dari *World Health Organization* (WHO), Penyakit Jantung merupakan penyebab utama

kematian di seluruh dunia [1]. Kementerian kesehatan Republik Indonesia juga menyampaikan bahwa penyakit jantung menjadi penyebab utama kematian di Indonesia [2]. Data dari Riset Kesehatan Dasar (Riskesdas) pada tahun 2013 dan 2018 menunjukkan bahwa prevalensi penyakit jantung mengalami peningkatan dari 0,5% pada tahun 2013 menjadi 1,5% pada tahun 2018 [3]. Di samping itu, 17,2% dari pasien yang mengalami gagal jantung di Indonesia meninggal selama perawatan di rumah sakit, dan 11,3% meninggal dalam satu tahun setelah perawatan. Selain itu, serta sebanyak 17% dari mereka harus menjalani rawat inap berulang karena gejala dan tanda gagal jantung yang memburuk [4].

Penyakit gagal jantung adalah salah satu jenis penyakit kardiovaskular (CVD) [5], proses identifikasi penyakit kardiovaskular (CVD) cukup sulit karena beberapa faktor kontributor yang berkontribusi pada CVD, seperti tekanan darah tinggi, kadar kolesterol, diabetes, denyut nadi yang tidak normal, dan banyak faktor lainnya [5]. Terkadang gejala CVD dapat bervariasi antar jenis kelamin. Sebagai contoh, pasien pria lebih cenderung mengalami nyeri dada, sementara pasien wanita memiliki gejala lain bersamaan dengan nyeri dada seperti ketidaknyamanan dada, mual, kelelahan ekstrem, dan sesak nafas [6]. Berbagai penelitian telah menyelidiki berbagai metode untuk memprediksi penyakit jantung, namun pada tahap awal, ketepatan prediksi terkendala oleh sejumlah faktor, termasuk kompleksitas model yang digunakan, waktu eksekusi yang diperlukan, dan akurasi hasil prediksi [7]. Maka dari itu, pengobatan dan diagnosis yang tepat dapat menyelamatkan banyak nyawa.

Diagnosis dini dapat membantu mencegah gagal jantung yang mengakibatkan kematian seseorang. Angiografi dianggap sebagai metode yang paling tepat dan akurat untuk memprediksi penyakit arteri koroner (CAD) [8], namun biayanya sangat tinggi sehingga membuatnya kurang dapat diakses oleh keluarga berpendapatan rendah. Biaya penyakit kardiovaskular memberikan dampak finansial yang besar pada sistem kesehatan Amerika Serikat. Pada periode tahun 2014 dan 2015, total biayanya mencapai sekitar \$219 miliar per tahun, mencakup layanan kesehatan, pengeluaran obat-obatan, dan produktivitas yang hilang akibat kematian [9].

Dalam era modern, catatan kesehatan elektronik (EHRs) juga bermanfaat untuk tujuan klinis dan penelitian [10]. *Data mining* memainkan peran besar dalam mengekstrak informasi yang berguna dari *big data*. Ini secara luas digunakan di hampir setiap bidang kehidupan seperti kedokteran, teknik, bisnis, dan pendidikan. Pertambahan data digunakan untuk menjelajahi data untuk mengekstrak informasi penting pengambilan keputusan dari kumpulan repositori masa lalu untuk masa depan. Berbagai algoritma pembelajaran mesin telah digunakan untuk memahami kompleksitas dan interaksi non-linear antara faktor-faktor yang berbeda dengan mengurangi kesalahan dalam prediksi dan hasil faktual [11]. Karena data medis yang terus meningkat, kita perlu memanfaatkan algoritma pembelajaran mesin untuk membantu profesional kesehatan medis dalam menganalisis data dan membuat keputusan diagnostik yang akurat dan tepat. Dalam

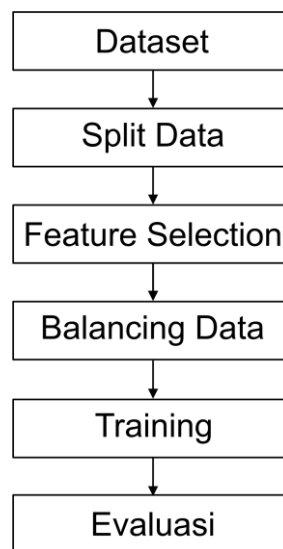
pertambahan data medis, berbagai algoritma klasifikasi digunakan untuk memprediksi CVD pada pasien dan prediksi kematian akibat serangan jantung [12].

Ahmad dkk. [13] merilis sebuah dataset yang terdiri dari catatan medis pasien jantung yang sebelumnya mengalami gagal jantung, yang dikumpulkan di *Institute of Cardiology dan Allied Hospital Faisalabad*, Pakistan. Para penulis memprediksi tingkat kematian dengan menerapkan regresi Cox. Penelitian tersebut juga menyoroti pola-pola kelangsungan hidup menggunakan Plot Kaplan-Meier. Penting untuk dicatat bahwa mereka telah membuat dataset tersebut tersedia untuk komunitas ilmiah. Ishaq, A. dkk [14] berhasil melakukan penerapan 9 algoritma *machine learning* (ML) terhadap dataset gagal jantung di Pakistan tersebut dan didapatkan akurasi terbaik hingga 92.6%. Namun pada penelitian tersebut hanya berfokus terhadap akurasi prediksi saja, sedangkan untuk data *binary* perlu diperhatikan metrik *f1-score* bahkan *Matthews correlation coefficient* (MCC) [15].

Dari hasil penelitian tersebut mendorong upaya untuk membantu profesional kesehatan dengan mengembangkan teknik pembelajaran mesin dalam diagnosis kelangsungan hidup pasien CVD. Penelitian ini akan dilakukan dengan menerapkan *ensemble techniques* [16], dengan model *Random Forest* [17], *Bagging* [18], *Adaboost* [19], *Gradient Boosting* [20] dan *XGBoost* [21]. Pada masing-masing model akan dilakukan penyetelan *hyperparameter* untuk mendapatkan model terbaik dalam memprediksi penyakit gagal jantung.

II. METODE PENELITIAN

Penelitian ini akan mengadopsi beberapa metode, metode tersebut ditunjukkan pada Gambar 1 berikut.



Gambar 1. Metode Penelitian

A. Dataset

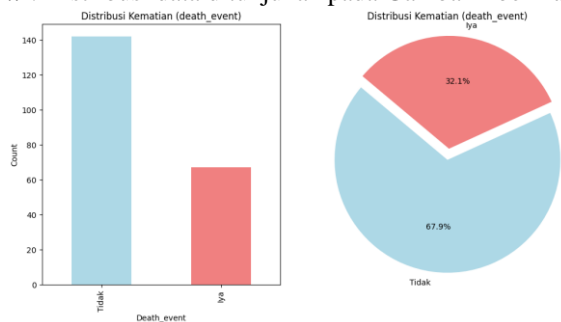
Penelitian ini menggunakan dataset yang dirilis oleh Ahmad dkk. [13] yang dikumpulkan berdasarkan analisis kelangsungan hidup pasien gagal jantung yang dirawat di *Institute of Cardiology dan Allied Hospital Faisalabad*

Pakistan selama bulan April-Desember (2015). Semua pasien berusia 40 tahun ke atas dan mengalami disfungsi sistolik ventrikel kiri, termasuk NYHA kelas III dan IV. Dataset yang terkumpul sebanyak 299 pasien gagal jantung yang terdiri dari 105 wanita dan 194 pria untuk detail spesifikasi dataset ditunjukkan pada Tabel I berikut.

TABEL I
SPESIFIKASI DATASET

No.	Atribut	Deskripsi	Nilai
1	Age	Usia pasien	40 - 95
2	Anemia	Status anemia	0 - 1
3	Creatinine Phosphokinase (CPK)	Kadar enzim CPK dalam darah	23 - 7861
4	Diabetes	Status diabetes	0 - 1
5	Ejection Fraction	Persentase ejeksi jantung	14 - 80
6	High Blood Pressure	Status darah tinggi	0 - 1
7	Platelets	Jumlah trombosit dalam darah	25100 - 850000
8	Serum Creatinine	Kadar kreatinin dalam serum	0.5 - 9.4
9	Serum Sodium	Kadar natrium dalam serum	113 - 148
10	Sex	Jenis kelamin pasien	0 - 1
11	Smoking	Status perokok	0 - 1
12	Time	Waktu pengamatan	4 - 285
13	Death Event (target)	Kejadian kematian	0 - 1

Dataset diatas memiliki distribusi data yang tidak seimbang berdasarkan atribut targetnya yaitu 'Death Event'. Distribusi data ditunjukkan pada Gambar 2 berikut.



Gambar 2. Distribusi Atribut Target

Maka dari itu diperlukan penanganan *balancing data* dalam proses selanjutnya.

B. Split Data

Tahap *split data* ini akan dilakukan proses pembagian dataset menjadi data latih dan data uji. Dengan memberikan rasio pembagian data dapat berpengaruh besar pada proses validasi pengujian model [22]. Rasio yang akan digunakan pada penelitian ini adalah 70:30 dengan detail dijelaskan pada Tabel II berikut.

TABEL II
PEMBAGIAN DATASET

No.	Data	Rasio	Jumlah Data
1	Data Latih	70 %	209
2	Data Uji	30 %	90

C. Feature Selection

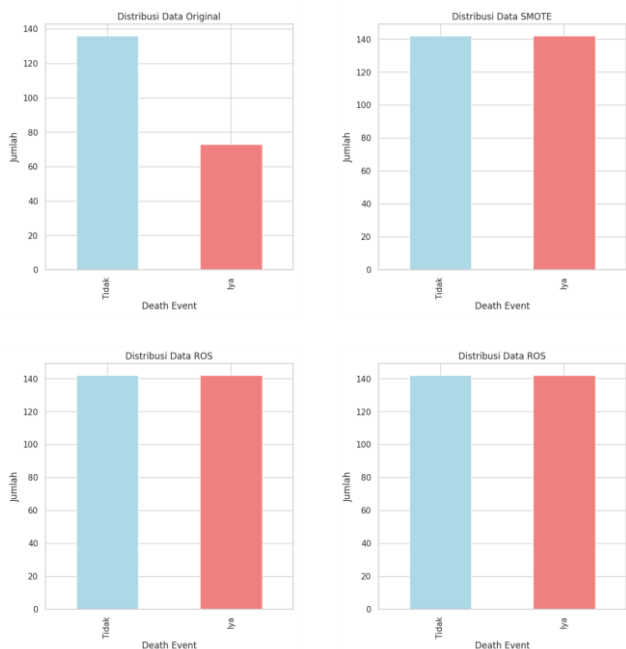


Gambar 3. Feature Importance

Random forest (RF) digunakan sebagai model untuk mendapatkan *feature importance* dikarenakan RF lebih unggul dibandingkan model lain [23]. Gambar 3 menunjukkan urutan pentingnya fitur, dan yang paling relevan adalah *time*, *serum_creatinine*, *ejection_fraction*, *age*, *CPK*, *serum_sodium*, *platelets*.

D. Balancing Data

Teknik *machine learning* modern berupaya mengatasi data yang tidak seimbang dengan cara melakukan *oversampling* atau *undersampling* [24]. SMOTE bekerja dengan cara membuat sampel sintesis (buatan) untuk kelas minoritas berdasarkan sampel yang sudah ada [25]. *Random Over-sampling* (ROS) adalah suatu metode dalam penanganan masalah ketidakseimbangan kelas pada pembelajaran mesin dengan meningkatkan jumlah sampel pada kelas minoritas dengan cara menambahkan beberapa salinan acak dari sampel yang sudah ada dalam kelas tersebut [26]. *Random Under-sampling* (RUS) metode lain dalam menangani masalah ketidakseimbangan kelas dengan mengurangi jumlah sampel pada kelas mayoritas dengan cara menghapus beberapa sampel secara acak. [27]. Hasil dari proses *balancing data* latih ditunjukkan pada Gambar 4 berikut.



Gambar 4. Balancing Data

E. Training

Hyperparameter tuning pada proses pelatihan model machine learning merupakan suatu yang hal penting dalam mendapatkan model ML yang efektif terhadap dataset [28]. Proses hyperparameter tuning dilakukan dengan menyetel nilai parameter pada setiap model. Pada penelitian ini melakukan hyperparameter tuning terhadap 5 model ensemble learning, detail parameter yang akan dituning ditunjukkan pada Tabel III berikut.

TABEL III
PARAMETER HYPERPARAMETER TUNING

No.	Model	Parameter	Nilai
1	Random Forest	n_estimators	[50, 100, 200]
		max_depth	[None, 10, 20, 30]
		min_samples_split	[2, 5, 10]
		min_samples_leaf	[1, 2, 4]
2	Bagging	base_estimator__max_depth	[None, 5, 10, 15]
		base_estimator__min_samples_split	[2, 5, 10]
		n_estimators	[10, 50, 100]
3	Adaboost	n_estimators	[50, 100, 200]
		learning_rate	[0.01, 0.1, 0.5, 1.0]
4	Gradient Boosting	n_estimators	[50, 100, 200]
		learning_rate	[0.01, 0.1, 0.5]
		max_depth	[3, 5, 7]
		min_samples_split	[2, 5, 10]
5	XGBoost	n_estimators	[50, 100, 200]
		learning_rate	[0.01, 0.1, 0.5]
		max_depth	[3, 5, 7]
		min_child_weight	[1, 3, 5]

	subsample	[0.8, 1.0]
	colsample_bytree	[0.8, 1.0]

Hasil hyperparameter tuning pada setiap modelnya ditampilkan pada Tabel IV berikut.

TABEL IV
HASIL HYPERPARAMETER TUNING

No.	Model	Parameter	Nilai
1	Random Forest	n_estimators	100
		max_depth	10
		min_samples_split	10
		min_samples_leaf	4
2	Bagging	base_estimator__max_depth	None
		base_estimator__min_samples_split	10
		n_estimators	100
3	Adaboost	n_estimators	50
		learning_rate	0.01
4	Gradient Boosting	n_estimators	200
		learning_rate	0.1
		max_depth	7
		min_samples_split	10
5	XGBoost	n_estimators	50
		learning_rate	0.1
		max_depth	5
		min_child_weight	3
		subsample	0.8
		colsample_bytree	0.8

F. Evaluasi

Evaluasi model dilakukan untuk mendapatkan model terbaik dengan melakukan pengujian model terhadap data uji. Evaluasi dilakukan dengan beberapa metrik yaitu accuracy, precision, recall, f1-score dan Matthews correlation coefficient (MCC).

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \tag{1}$$

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \tag{2}$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \tag{3}$$

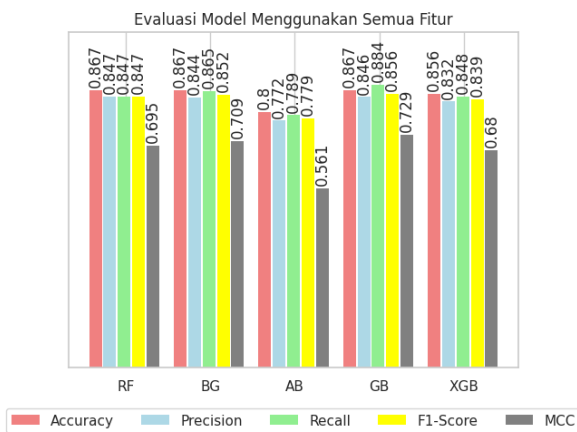
$$F1\ -\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

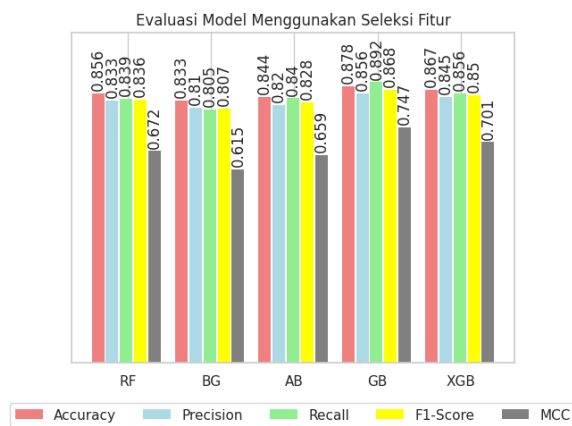
III. HASIL DAN PEMBAHASAN

Eksperimen dilakukan untuk mendapatkan model *ensemble* dengan evaluasi terbaik untuk prediksi kelangsungan hidup pasien dengan gagal jantung. Model akan diuji terhadap data uji, baik dengan data semua fitur atau dengan fitur terpenting, dan model dengan hasil *balancing data*.

A. Hasil Eksperimen Model Original



Gambar 5. Hasil Eksperimen Semua Fitur



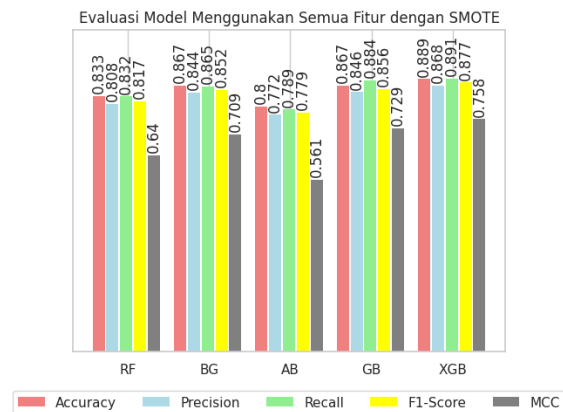
Gambar 6. Hasil Eksperimen Menggunakan 7 Fitur

Gambar 5 dan 6 menampilkan hasil kinerja model terhadap dataset dengan semua fitur dan dataset dengan 7 fitur. Dataset pada eksperimen ini masih imbalance, dan diketahui hasil evaluasi dalam kedua situasi, algoritma *Gradient Boosting* (GB) tetap menjadi pilihan utama dengan performa yang konsisten dan optimal. Pada dataset dengan seluruh fitur, GB menunjukkan akurasi sebesar 86.7%, *F1-score* sebesar 85.6%, dan *Matthews Correlation Coefficient* (MCC) sebesar 72.9%, sementara pada dataset dengan seleksi fitur, GB mencapai akurasi 87.8%, *F1-score* 86.8%, dan MCC 74.7%. Temuan ini menunjukkan bahwa GB mampu mengatasi perbedaan dalam jumlah fitur, dan bahkan dengan fitur yang lebih sedikit, performanya tetap tinggi.

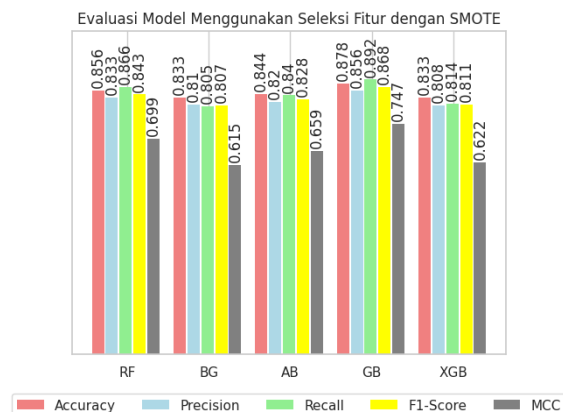
Pentingnya penerapan seleksi fitur juga diperkuat oleh penurunan performa yang diamati pada beberapa algoritma, seperti *Random Forest* (RF) dan *Bagging* (BG), setelah menggunakan dataset dengan seleksi fitur. Ini menunjukkan bahwa tidak semua algoritma memiliki fleksibilitas yang sama terhadap perubahan dalam jumlah fitur, dan dalam beberapa kasus, pengurangan fitur dapat mengurangi kemampuan prediktif algoritma. Namun demikian, *Extreme Gradient Boosting* (XGB) menunjukkan performa yang konsisten di kedua kondisi dataset, menegaskan bahwa tidak semua algoritma merespon seleksi fitur dengan cara yang sama.

Selain itu, penurunan performa yang signifikan pada algoritma *AdaBoost* (AB) di kedua situasi dataset menyoroti kelemahan tertentu dalam algoritma tersebut, yang perlu dipertimbangkan saat memilih model untuk dataset yang tidak seimbang.

B. Hasil Eksperimen Model SMOTE



Gambar 7. Hasil Eksperimen Semua Fitur dengan SMOTE



Gambar 8. Hasil Eksperimen Menggunakan 7 Fitur dengan SMOTE

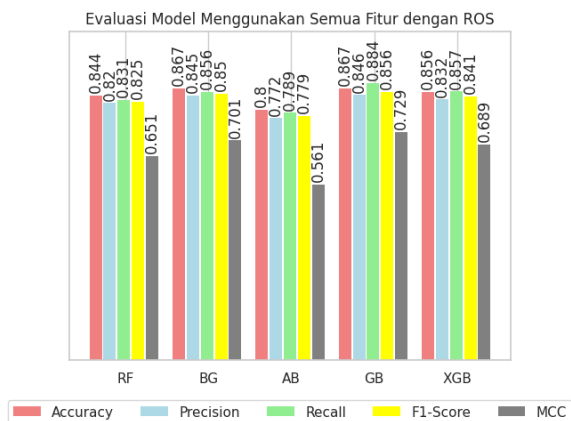
Dari hasil eksperimen dan analisis yang ditunjukkan pada Gambar 7 dan 8, terlihat bahwa penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) memberikan dampak yang signifikan terhadap performa algoritma klasifikasi pada kedua kondisi dataset, baik pada dataset dengan seluruh fitur maupun dataset setelah dilakukan seleksi fitur. Langkah ini menjadi penting karena ketidakseimbangan kelas dapat menyebabkan bias dalam model klasifikasi, di mana model cenderung lebih

baik dalam memprediksi kelas mayoritas daripada kelas minoritas.

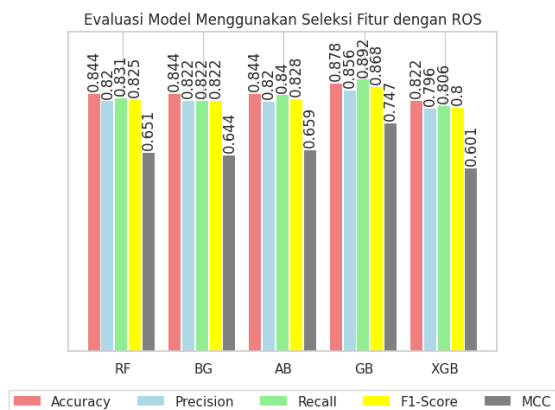
Pada dataset dengan seluruh fitur, terlihat bahwa algoritma *Extreme Gradient Boosting* (XGB) mencapai akurasi tertinggi sebesar 88.9%, yang disertai dengan nilai presisi, *recall*, *F1-score*, dan *Matthews Correlation Coefficient* (MCC) yang sangat baik. Hasil ini menunjukkan bahwa XGB mampu menangani perubahan dalam fitur dengan baik dan memberikan prediksi yang konsisten. Namun, pada dataset seleksi fitur, meskipun XGB mengalami penurunan performa, algoritma *Gradient Boosting* (GB) tetap mendominasi dengan akurasi sebesar 87.8% dan mencapai nilai *F1-score* dan MCC yang tinggi.

Selain itu, penerapan SMOTE pada dataset seleksi fitur juga berhasil meningkatkan performa *Random Forest* (RF) dan *AdaBoost* (AB), menunjukkan efektivitas SMOTE dalam meningkatkan keberlanjutan algoritma klasifikasi pada dataset yang mengalami ketidakseimbangan kelas. Namun, terdapat beberapa algoritma, seperti *Bagging* (BG) dan XGB, yang mengalami penurunan performa pada dataset seleksi fitur. Hal ini menandakan bahwa efek SMOTE tidak selalu konsisten di semua algoritma dan dapat bergantung pada karakteristik data dan algoritma yang digunakan.

C. Hasil Eksperimen Model ROS



Gambar 9. Hasil Eksperimen Semua Fitur dengan ROS



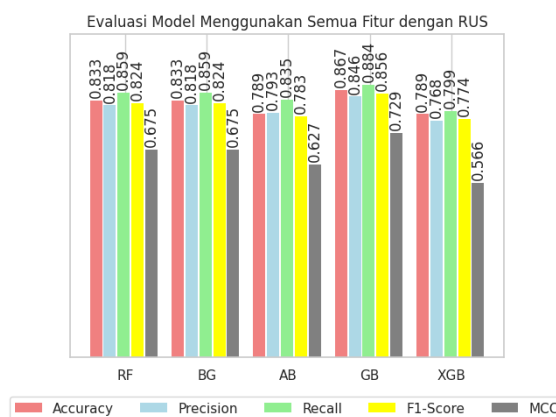
Gambar 10. Hasil Eksperimen Menggunakan 7 Fitur dengan ROS

Pada eksperimen ini, penerapan teknik *Random Over Sampling* (ROS) pada dataset dengan seluruh fitur dan dataset setelah dilakukan seleksi fitur telah memberikan

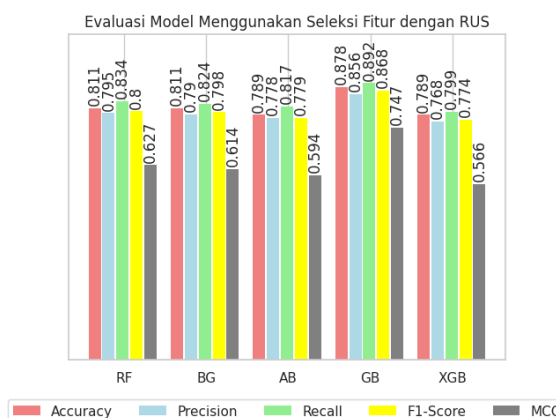
pengaruh dari ROS terhadap performa model klasifikasi. Sebagaimana ditunjukkan pada Gambar 9 dan 10, hasilnya menunjukkan bahwa ROS berhasil meningkatkan performa beberapa algoritma pada dataset dengan seluruh fitur. Algoritma *Gradient Boosting* (GB) mencapai akurasi tertinggi sebesar 86.7%, dengan nilai presisi, *recall*, *F1-score*, dan *Matthews Correlation Coefficient* (MCC) yang baik. Pada dataset seleksi fitur setelah ROS, GB tetap mendominasi dengan akurasi 87.8% dan nilai *F1-score* serta MCC yang tinggi.

Namun, perlu diperhatikan bahwa beberapa algoritma, seperti *Random Forest* (RF) dan *Bagging* (BG), mengalami penurunan performa pada dataset seleksi fitur setelah ROS, sementara *Extreme Gradient Boosting* (XGB) menunjukkan penurunan yang lebih signifikan. Hal ini menunjukkan bahwa pengaruh ROS tidak konsisten di semua algoritma dan dapat bergantung pada karakteristik dataset, terutama setelah dilakukan seleksi fitur.

D. Hasil Eksperimen Model RUS



Gambar 11. Hasil Eksperimen Semua Fitur dengan RUS



Gambar 12. Hasil Eksperimen Menggunakan 7 Fitur dengan RUS

Dari hasil evaluasi eksperimen terakhir seperti pada Gambar 11 dan 12, terlihat bahwa penerapan teknik *Random Under Sampling* (RUS) memberikan dampak yang signifikan terhadap performa algoritma klasifikasi pada kedua jenis dataset, baik dengan seluruh fitur maupun setelah dilakukan seleksi fitur. Secara khusus, pada dataset dengan seluruh fitur, algoritma *Gradient Boosting* (GB) menunjukkan performa paling unggul dengan mencapai akurasi tertinggi sebesar 86.7%, serta

nilai presisi, *recall*, *F1-score*, dan *Matthews Correlation Coefficient* (MCC) yang tinggi. Bahkan, setelah RUS diterapkan pada dataset dengan seleksi fitur, GB mengalami peningkatan akurasi mencapai 87.8%, serta mencapai nilai tertinggi dalam *F1-score* dan *MCC*. Meskipun demikian, penurunan performa terlihat pada beberapa algoritma seperti *Random Forest* (RF) dan *Bagging* (BG) setelah dilakukan seleksi fitur dan penerapan RUS, namun algoritma *Adaboost* (AB) dan *Extreme Gradient Boosting* (XGB) menunjukkan performa yang stabil dengan hasil evaluasi terendah.

Peningkatan yang signifikan dalam performa GB setelah RUS menunjukkan efektivitas teknik tersebut dalam menangani ketidakseimbangan kelas pada dataset. Meskipun demikian, perubahan performa yang bervariasi pada algoritma lain menunjukkan bahwa dampak dari RUS dapat bergantung pada karakteristik masing-masing algoritma dan dataset.

IV. KESIMPULAN

Dalam eksperimen memprediksi kematian pada pasien gagal jantung menggunakan dataset *heart failure*, penelitian ini mengevaluasi performa lima algoritma klasifikasi, yaitu *Random Forest* (RF), *Bagging* (BG), *AdaBoost* (AB), *Gradient Boosting* (GB), dan *Extreme Gradient Boosting* (XGB), pada berbagai kondisi dataset. Meskipun hasil eksperimen menunjukkan bahwa mayoritas hasil evaluasi tertinggi didapatkan oleh model *Gradient Boosting*, namun model terbaik diperoleh dengan menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada algoritma *Extreme Gradient Boosting* (XGB) pada dataset dengan seluruh fitur, mencapai akurasi tertinggi sebesar 88.9%, *F1-score* 87.7%, dan *Matthews Correlation Coefficient* (MCC) 75.8%. Penerapan metode *balancing data*, seperti SMOTE, ROS, dan RUS, memberikan dampak yang signifikan terhadap performa algoritma, namun perlu dipilih dengan cermat sesuai dengan karakteristik dataset. Dengan demikian, temuan ini diharapkan memberikan kontribusi penting dalam upaya meningkatkan prediksi dan manajemen risiko kematian lebih awal pada pasien gagal jantung dalam bidang kesehatan.

DAFTAR PUSTAKA

- [1] WHO, "The top 10 causes of death." <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] Rokom, "Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer." <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220929/0541166/penyakit-jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer/>
- [3] Kementretian Kesehatan RI, "Hasil Utama Riskesdas 2018," 2018. <https://kesmas.kemkes.go.id/> (accessed Apr. 28, 2024).
- [4] R. Lestari, "Angka Kematian karena Gagal Jantung di Indonesia Tergolong Tinggi," 2022. <https://www.medcom.id/gaya/fitness-health/0k8XZ4gk-angka-kematian-karena-gagal-jantung-di-indonesia-tergolong-tinggi> (accessed Apr. 28, 2024).
- [5] Cooper *et al.*, "Trends and disparities in coronary heart disease, stroke, and other cardiovascular diseases in the United States: findings of the national conference on cardiovascular disease prevention," *Circulation*, vol. 102, no. 25, pp. 3137–3147, 2000.
- [6] S. Suman, J. Pravalika, P. Manjula, and U. Farooq, "Gender and CVD- Does It Really Matters?," *Curr. Probl. Cardiol.*, vol. 48, no. 5, 2023.
- [7] L. A. Allen *et al.*, "Decision Making in Advanced Heart Failure A Scientific Statement From the American Heart Association," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.
- [8] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Y. A., "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, 2017.
- [9] C. T. L. Emelia J Benjamin, Paul Muntner, Alvaro Alonso, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Sandeep R Das, Francesca N Delling, Luc Djousse, Mitchell S V Elkind, Jane F Ferguson, Myriam F *et al.*, "Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association," *Circulation*, vol. 139, no. 10, pp. 356–e528, 2019.
- [10] B. Chapman, A. D. DeVore, R. J. Mentz, and M. Metra, "Clinical profiles in acute heart failure: an urgent need for a new approach," *ESC Hear. Fail.*, vol. 6, no. 3, pp. 464–474, 2019.
- [11] S. F. Weng, J. Reys, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLoS One*, vol. 12, no. 4, p. e0174944, 2017.
- [12] V. V Ramalingam and A. Dandapath, "Heart disease prediction using machine learning techniques: A survey," *Int. J. Eng. Technol.*, vol. 7, no. 2.8, p. 684, 2017.
- [13] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," *PLoS One*, vol. 12, no. 7, p. e0181001, 2017.
- [14] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021.
- [15] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 6, pp. 1–13, 2020.
- [16] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," *J. Hydrol.*, vol. 598, no. 33, 2021.
- [17] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.
- [18] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.
- [19] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Stat. Interface*, vol. 2, no. 3, pp. 349–360, 2009.
- [20] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurobot.*, vol. 7, p. 21, 2013.
- [21] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proc. 22nd acm sigkdd Int. Conf. Knowl. Discov. data Min.*, pp. 785–794, 2016.
- [22] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification," *Molecules*, vol. 26, no. 4, p. 1111, 2021.
- [23] A. Altmann, L. Toloși, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.
- [24] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci. (Ny)*, vol. 513, pp. 429–441, 2020.
- [25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [26] M. Hayaty, S. Muthmainah, and S. M. Ghufuran, "Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification," *Int. J. Artif. Intell. Res.*, vol. 4, no. 20, pp. 86–94, 2020.
- [27] T. Hasanin and T. Khoshgoftaar, "The Effects of Random Undersampling with Simulated Class Imbalance for Big Data," in *IEEE International Conference on Information Reuse and*

- [28] *Integration (IRI)*, 2018, pp. 70–79.
S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil,
and S. Patil, “Liver Diseases Prediction using KNN with
Hyper Parameter Tuning Techniques,” in *2020 IEEE
Bangalore Humanitarian Technology Conference (B-HTC)*,
2020, pp. 1–6. doi: 10.1109/B-HTC50970.2020.9297949.