Metode *Weighted Random Forest* dalam Klasifikasi Prediksi Kelangsungan Hidup Pasien Gagal Jantung

Laila Budianti*, Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*laila.budianti28@gmail.com, suliadi@gmail.com

Abstract. Random forest classification is an analysis of a data class model in predicting the class of an unknown object through the majority vote. It has high performance in classifying object with balanced classes, but it does not so good in the case of imbalanced data, where the resulting predictions tend to be biased to the majority class. Recent approach for imbalanced data classification is Weighted Random Forest introduced by Chen and Breiman (2004). This method can solve the minority class prediction problem and improve accuracy by giving specific weight to each classes. This method can also be applied to medical data. In the survival of the heart failure patient belongs to the imbalanced data, where the data on heart failure patients that survived are much more than heart failure patients that died. In this research, we use Weighted Random Forest to classify heart failure patients in RSUP Dr. Wahidin sudirohusodo Makassar. The data used in this research are obtained from Nugraha (2017) the result showed that Weighted Random Forest classification methods are better than random forest because they can increase the accuracy by 1.61%, f-measure by 0.92%, and the AUC (area under the curve) by 0.89%. The result of the evaluation Weighted Random Forest classification for accuracy is 90.32%, precision is 93.22%, recall is 96.49%, f-measure is 94.82%, and AUC score is 0.5825.

Keywords: *Heart Failure, Classification, Weighted Random Forest, Imbalanced Data.*

Abstrak. Klasifikasi *random forest* merupakan analisis untuk menemukan model kelas data dalam memprediksi kelas dari objek yang belum diketahui datanya melalui suara mayoritas. Metode ini memiliki kinerja yang baik dalam mengklasifikasikan data dengan kelas seimbang, tetapi tidak baik dalam kasus *imbalanced data*, di mana prediksi yang dihasilkan cenderung bias terhadap kelas mayoritas. Pendekatan terbaru untuk klasifikasi *imbalanced data* adalah *Weighted Random Forest* yang diperkenalkan oleh Chen dan Breiman (2004). Metode ini dapat menyelesaikan masalah prediksi kelas minoritas dan meningkatkan akurasi dengan memberikan bobot pada setiap kelas. Metode ini juga dapat diterapkan pada data medis. Dalam kelangsungan hidup pasien gagal jantung tergolong *imbalanced data*, dimana data pasien gagal jantung yang selamat jauh lebih banyak dibandingkan pasien gagal jantung yang meninggal. Dalam penelitian ini, kami menggunakan *Weighted Random Forest* untuk mengklasifikasikan pasien gagal jantung di RSUP Dr. Wahidin Sudirohusodo Makassar. Data yang digunakan dalam penelitian ini diperoleh dari Nugraha (2017) hasil penelitian menunjukkan bahwa metode klasifikasi *Weighted Random Forest* lebih baik daripada *random forest* karena dapat meningkatkan accuracy sebesar 1,61%, *F-Measure* sebesar 0,92%, dan nilai AUC (*area under the curve*) sebesar 0,89%. Hasil evaluasi untuk klasifikasi *Weighted Random Forest* untuk *accuracy* sebesar 90,32%, *precision* sebesar 93,22%, *recall* sebesar 96,49%, *F-Measure* sebesar 94,82%, dan nilai AUC sebesar 0,5825.

Kata Kunci: *Gagal Jantung, Klasifikasi, Weigthed Random Forest, Imbalanced Data.*

Corresponding Author
Email: suliadi@gmail.com

A. Pendahuluan

Klasifikasi dalam data *mining* adalah teknik untuk memisahkan data ke dalam kelas berbeda (Han *et al.*, 2012). Salah satu algoritma klasifikasi adalah *random forest*. Klasifikasi dengan *random forest* dilakukan dengan membangun beberapa pohon keputusan yang menghasilkan satu suara prediksi dari suara mayoritas (Xu *et al.*, 2020). Pohon keputusan dibangun dengan menggunakan *Classification and Regression Trees* (CART). *Random forest* dapat menghasilkan akurasi yang tinggi dibandingkan dengan metode klasifikasi lainnya. Hanya saja saat *random forest* dijalankan pada *imbalanced data* akan mengakibatkan kemampuan klasifikasi menurun. Dalam algoritma pengklasifikasian termasuk *random forest* menganggap bahwa kelas minoritas ini sebagai *noise* atau *outlier* yang akan diabaikan, sehingga hasil prediksi klasifikasi cenderung pada kelas mayoritas. Akibatnya akurasi dari prediksi untuk kelas minoritas akan jauh lebih kecil dibandingkan kelas mayoritas. Untuk mengatasi hal tersebut, Chen dan Breiman (2004) melakukan modifikasi pada algoritma *random forest* dengan memberikan bobot pada setiap kelas yang disebut *Weighted Random Forest*. Pembobotan dilakukan untuk setiap kelas dengan bobot tertinggi diberikan pada kelas minoritas.

Klasifikasi pada bidang medis dalam praktiknya merupakan kasus *imbalanced data* yang dimana frekuensi kelas tidak seimbang. *Imbalanced data* adalah suatu keadaan dimana frekuensi kelas data tidak seimbang (Siringoringo, 2018). Sehingga dalam melakukan klasifikasi dengan *imbalanced data* dapat menggunakan *Weighted Random Forest*.

Gagal jantung merupakan penyakit kardiovaskular sebagai stadium akhir dari semua penyakit jantung. Berdasarkan data WHO tahun 2022 menyebutkan sekitar 17,9 juta orang di dunia meninggal karena penyakit jantung. Kementrian Kesehatan pada tahun 2019 memperkirakan pada tahun 2020 ada sekitar 24,6% kasus kematian dikarenakan penyakit jantung. Dengan pravelensi pasien gagal jantung di Indonesia sebesar 5% dari jumlah populasi lebih tinggi dibandingkan data pravelensi populasi eropa yang hanya berkisar 1%-2%. Dengan keadaan yang demikian, maka diperlukan perhatian yang lebih dalam mengidentifikasi dini masalah pasien yang mengalami gagal jantung.

Berdasarkan latar belakang yang telah diuraikan, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Untuk mengetahui metode klasifikasi *Weighted Random Forest* dalam memperbaiki prediksi dari metode *random forest* standar.
2. Untuk mengetahui hasil perbandingan metode klasifikasi antara *Weighted Random Forest* dan *random forest* standar dalam klasifikasi pasien gagal jantung.

B. Metodologi Penelitian

Classification and Regression Trees (CART)

CART merupakan algoritma *decision tree*. *Decision tree* dapat diterapkan untuk mempelajari klasifikasi dan memprediksi pola dari data yang menggambarkan relasi dari variabel prediktor dan variabel respon dalam bentuk pohon. Menurut Timofeev (2004) tujuan utama dari metode CART adalah untuk mendapatkan kumpulan data yang akurat sebagai fitur dari suatu pengklasifikasian. Pohon klasifikasi terbentuk dari proses pemilahan rekursif biner pada suatu gugus data sehingga nilai variabel respon pada setiap gugus dan hasil pemilahan akan lebih homogen.

Tahapan pembentukan pohon klasifikasi terdiri dari:

1. Penentuan pemilah untuk variabel prediktor X_p adalah nominal maka M nilai tak berurut yang berbeda memiliki $2^{M-1}-1$ pemilah. Sedangkan untuk variabel prediktor adalah ordinal dan kontinu maka M nilai berbeda adalah $M-1$. Fungsi *impurity* adalah indeks gini. Indeks gini merupakan pengukuran tingkat keragaman suatu kelas dari suatu simpul tertentu dalam klasifikasi sehingga mampu membantu dalam menemukan fungsi pemilah yang optimal. Indeks gini berdasarkan rumus:

$$I(t) = \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} P(j|t)P(j'|t) = 1 - \sum_{j=1}^{J} P^2(j|t)$$

di mana $I(t)$ adalah indeks gini pada simpul t dan $P(j|t)$ adalah probabilitas kelas j dalam sebuah simpul t dengan $P(j|t) = \frac{n_j(t)}{n(t)}$. Setelah dilakukan pemilahan semua kemungkinan, selanjutnya menentukan kriteria *goodness of split* berdasarkan rumus:

$\phi(s,t) = \Delta i(s,t) = i(t) - P_L i(t_L) - P_R i(t_R)$

Pemilah yang terpilih untuk simpul t jika memaksimalkan goodness of split dan memberikan penurunan keheterogenan tertinggi, berdasarkan rumus:

$\Delta i(s^*, t_i) = max_{s \epsilon S} \Delta i(s,t)$

2. Penandaan label kelas pada simpul t ditentukan berdasarkan pada proporsi kelas terbesar berdasarkan rumus:

$$\boldsymbol{P(j_0, t) = max_j P(j|t) = max_j \frac{n_j(t)}{n(t)}}$$

di mana $P(j|t)$ adalah proporsi kelas-j pada simpul t, $N_j(t)$ adalah banyaknya amatan kelas ke-j pada simpul ke-t, dan $N\ (t)$ adalah banyaknya pengamatan pada simpul t.

3. Penentuan simpul terminal jika, hanya terdapat sebuah nilai variabel pada respon, saat dilakukan pemilahan tidak terdapat penurunan tingkat keragaman dan simpul hanya terdapat satu kasus.

Random Forest

Random forest merupakan metode *ensemble* yang digunakan untuk mengatasi masalah klasifikasi. Metode *ensemble* merupakan cara untuk meningkatkan akurasi model klasifikasi dengan mengombinasikan metode klasifikasi (Han *et al.*, 2012). *Random forest* dapat memberikan hasil klasifikasi yang sangat baik dengan hasil residu/*error* yang rendah, dapat menangani data *training* yang sangat besar secara efisien, dan metode *random forest* efektif dalam mengatasi masalah *missing data* (Breiman, 2001).

Menurut Cutler et al. (2011), *random forest* merupakan pengembangan metode CART dengan menerapkan metode *Boostrap Aggregating* (*Bagging*) dan *Random Feature Selection* (8). Di dalam *random forest* setiap pohon keputusan berisi kumpulan variabel acak. Dalam pengertian statistik, X_1, X_2, … ,X_p di mana merupakan variabel prediktor dan Y merupakan variabel respon, di mana X dan Y yaitu variabel acak yang diambil bersama nilainya dari $\mathcal{X} \times \mathcal{y}$. *Random forest* adalah pengklasifikasi yang terdiri dari kumpulan pengklasifikasi berstruktur pohon $\{h(X, \Theta_k), k = 1, \ldots, K\}$ di mana k adalah banyaknya pohon yang dibangun $\{\Theta_k\}$ adalah vektor acak independen yang terdistribusi identik.

Algoritma *random forest* menurut Breiman (2001) pada klasifikasi adalah sebagai berikut:

1. Membagi data menjadi 2 yaitu data *training* untuk melakukan pemodelan dengan proporsi sebesar 75% dari data asli dan data *testing* untuk evaluasi model dengan proporsi sebesar 25% dari data asli.
2. Tahap *bootstrap sample* untuk memulai membangun *random forest*, langkah pertama yang dilakukan adalah pengambilan *random sampling* berukuran n dari kumpulan data *training* dengan pengembalian berdasarkan rumus $\mathcal{D} = \{(X_1, Y_1), \cdots, (X_N, Y_N)\}$, dengan $X_i \epsilon \mathcal{X}$ dan di mana $X_i = (X_{i,1}, \cdots, X_{i,p})$.
3. Dengan hasil *bootstrap*, lakukan CART dengan membangun pohon keputusan sampai mencapai ukuran maksimum. Pada setiap simpul, jumlah variabel yang diambil berdasarkan rumus $mtry = \sqrt{p}$.
4. Ulangi langkah 2 dan 3 sebanyak k kali untuk menghasilkan hutan yang terdiri dari k pohon.
5. Lakukan penggabungan berdasarkan k buah pohon menggunakan pemungutan suara terbanyak (*majority vote*) yaitu $f(x) = \arg max_Y \sum_{k=1}^{K} I(h_k(x) = Y)$.

Weighted Random Forest

Klasifikasi *random forest* yang cenderung bias terhadap kelas mayoritas, maka akan mempengaruhi klasifikasi pada kelas minoritas. Untuk mengatasi hal ini, setiap kelas dapat ditetapkan bobot (*weight*), memberikan bobot yang lebih besar kepada kelas minoritas daripada

kelas mayoritas (Wong & Yeh, 2019). Misalkan c_p adalah biaya kesalahan klasifikasi dari kelas minoritas (positif) dan c_n adalah biaya kesalahan dari klasifikasi kelas mayoritas (negatif). Dalam prosedur induksi pohon, bobot kelas digunakan untuk menimbang kriteria gini dalam menemukan *split*. Pada simpul terminal setiap pohon, bobot kelas dipertimbangkan. Misalkan w_p adalah bobot biaya kesalahan klasifikasi dari kelas minoritas dan w_n adalah bobot biaya kesalahan klasifikasi dari kelas mayoritas (Wong & Yeh, 2019). Pembobotan kelas minoritas $w_p = \frac{c_p}{c_p+c_n}$, pembobotan kasus mayoritas $w_n = \frac{c_n}{c_p+c_n}$, dan pembobotan indeks gini $i(t) = 1 - w_p P_p^2 - w_n P_n^2$. Prediksi kelas dari setiap simpul terminal ditentukan oleh suara mayoritas tertimbang (*weighted majority vote*) yaitu, suara tertimbang dari suatu kelas adalah bobot untuk kelas tersebut dikalikan jumlah kasus untuk kelas tersebut pada simpul terminal, $f(x) = \arg max_Y \sum_{k=1}^{K} w\, I(h_k(x) = Y)$.

Evaluasi Kinerja Klasifikasi

Menurut Gorunescu (2010) *confusion matrix* adalah tabulasi dari perhitungan evaluasi kinerja model klasifikasi berdasarkan jumlah objek penelitian yang diprediksi dengan benar dan salah. *Confusion matrix* dapat digunakan untuk menilai kualitas *classifier* yaitu seberapa baik *classifier* dapat mengenali fitur dari kelas yang berbeda yang direpresentasikan dalam bentuk tabulasi silang, di mana kelas data yang diprediksi ditampilkan sebagai kolom matriks dan kelas data aktual ditampilkan sebagai baris matriks (Han *et al*., 2012).

Tabel 1. Confusion Matrix

Classification		***Predict Class***	
		Positive	***Negative***
Actual Class	***Positive***	***True Positive* (TP)**	***False Negative (FN)***
	Negative	***False Positive (FP)***	***True Negative (TN)***

Dari *confusion matrix* dapat memperoleh ukuran evaluasi kinerja klasifikasi yaitu *accuracy, precision, recall, F-measure,* dan nilai AUC (*area under curve*).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

$$\text{F} - \text{measure} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{TN+FP}$$

$$AUC = \frac{1+TPR-FPR}{2}$$

Data

Data yang digunakan adalah data sekunder yang diperoleh dari penelitian Nugraha (2017). Data tersebut berisikan catatan medis 245 pasien yang telah terdiagnosis gagal jantung di rawat inap RSUP Dr. Wahidin Sudirohusodo Makassar. Data medis pasien dicatat selama 1 Januari – 30 September 2017. Variabel yang digunakan dalam penelitian ini adalah:

1. Status akhir pasien (Y) : 0 (selamat), 1 (meninggal)
2. Umur pasien (X_1) dalam tahun
3. Jenis kelamin (X_2) : 0 (laki-laki), 1 (perempuan)
4. Riwayat hipertensi pasien (X_3) : 0 (tidak), 1 (ya)
5. Riwayat diabetes melitus pasien (X_4) : 0 (tidak), 1 (ya)
6. Merokok (X_5) : 0 (tidak merokok), 1 (merokok)

C. Hasil Penelitian dan Pembahasan

Klasifikasi dengan Metode *Random Forest*

Dalam penelitian ini banyaknya variabel prediktor yang digunakan untuk setiap pemilah ada sebanyak 2, dengan menggunakan 10-*fold cross validation* untuk memilih model terbaik dengan k=100, 250, 500, dan 1000 pohon. Hasil dari 10-*fold cross validation* menunjukkan bahwa model klasifikasi terbaik jika menggunakan nilai $k = 100$ dengan nilai OOB error sebesar 20,22%. Evaluasi model dengan melakukakan prediksi terhadap data testing menggunakan metode *random forest* disajikan pada Tabel 2.

Tabel 2. Confusion Matrix Metode Random Forest

Classification		*Predict Class*		
		Selamat	**Meninggal**	**Total**
Actual Class	**Selamat**	54	3	57
	Meninggal	4	1	5
	Total	58	4	62

Berdasarkan tabel 3 diperoleh nilai evaluasi model klasifikasi *random forest* yaitu *accuracy* sebesar 88,71%, *precision* sebesar 93,10%, r*ecall* sebesar 94,74%, *F-Measure* sebesar 93,90%, dan nilai AUC sebesar 0,5737.

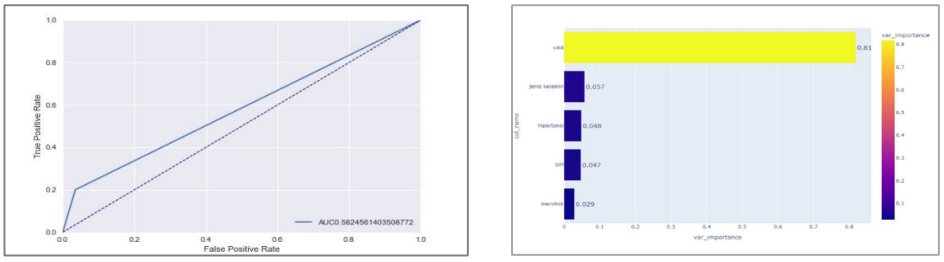
Klasifikasi dengan Metode *Weighted Random Forest*

Dalam klasifikasi *Weighted Random Forest* adalah menentukan banyaknya variabel prediktor, banyaknya pohon, bobot untuk kelas minoritas dan mayoritas. Banyaknya variabel prediktor yang digunakan untuk setiap pemilah ada sebanyak 2, dengan menggunakan 10-*fold cross validation* untuk memilih model terbaik dengan k=100, 250, 500, dan 1000 pohon. Hasil dari 10-$fold\ cross\ validation$ menunjukkan bahwa model klasifikasi terbaik jika menggunakan nilai $k = 100$. Menentukan bobot untuk setiap kelas berdasarkan biaya kesalahan klasifikasi minoritas dan mayoritas. Untuk biaya kesalahan minoritas c_p=4 dan biaya kesalahan mayoritas c_n=3. Sehingga bobot kelas minoritas sebesar 0,57 dan bobot kelas mayoritas sebesar 0,43. Dengan nilai OOB *error* sebesar 20,76%. Evaluasi model dengan melakukan prediksi terhadap data *testing* menggunakan metode *Weighted Random Forest* disajikan pada Tabel 3.

Tabel 3. Confusion Matrix Metode *Weighted Random Forest*

Classification		*Predict Class*		
		Selamat	**Meninggal**	**Total**
Actual Class	**Selamat**	55	2	57
	Meninggal	4	1	5
	Total	59	3	62

Dari nilai evaluasi model klasifikasi *Weighted Random Forest* didapatkan bahwa *accuracy* sebesar 90,32%, *precision* sebesar 93,22%, *recall* sebesar 96,49%, *F-Measure* sebesar 94,82%.



Gambar 1. Kurva ROC dan Variabel *importance* Metode *Weighted Random Forest*

Pada Gambar 1 menunjukkan kurva ROC dan variabel *importance* untuk klasifikasi dengan menggunakan metode *Weighted Random Forest*. Kurva ROC yang menggambarkan hubungan antara data *testing* dan data prediksi dengan nilai AUC yang diperoleh sebesar 0,5825 yang berarti hasil klasifikasi yang didapatkan masih lemah. Hal ini dikarenakan data *testing* pada kelas meninggal kecil sehingga nilai FPR menghasilkan nilai yang besar dan mengakibatkan nilai AUC lemah dalam menghasilkan klasifikasi. Namun berdasarkan nilai evaluasi klasifikasi yang telah didapatkan di atas, klasifikasi dengan metode *Weighted Random Forest* sudah baik dalam memprediksikan kelas pasien gagal jantung dengan mempertimbangkan nilai *accuracy* dan *F-Measure*. Sedangkan, variabel *importance* metode *Weighted Random Forest* menujukkan bahwa variabel usia merupakan variabel yang sangat penting terhadap prediksi kelangsungan hidup pasien gagal jantung pada metode *Weighted Random Forest* dilihat dari nilai variabel *importance* yang paling besar yaitu 0,81 dan memiliki nilai gini yang paling besar dari variabel yang lain.

Perbandingan Kinerja Metode Klasifikasi

Perbandingan kinerja metode klasifikasi dilakukan berdasarkan nilai evaluasi kinerja yang di dapat untuk mengetahui metode mana yang baik dalam melakukan klasifikasi pasien gagal jantung.

Tabel 4. Perbandingan Metode Random Forest dan *Weighted Random Forest*

Ukuran	*Random Forest*	*Weighted Random Forest*
Accuracy	88,71%	90,32%
Precision	93,10%	93,22%
Recall	94,74%	96,49%
F-Measure	93,90%	94,82%
AUC	57,36%	58,25%

Berdasarkan Tabel 4 dapat dilihat nilai evaluasi kinerja klasifikasi metode *random forest* dan *Weighted Random Forest*. Metode *Weighted Random Forest* mendapatkan nilai evaluasi kinerja klasifikasi lebih tinggi dibandingkan dengan metode *random forest*. Klasifikasi *Weighted Random Forest* untuk *accuracy* mengalami kenaikan sebesar 1,61%, *precision* mengalami kenaikan sebesar 0,12%, *recall* mengalami kenaikan sebesar 1,74%, *F-Measure* mengalami kenaikan sebesar 0,92%, dan nilai AUC mengalami kenaikan sebesar 0,89% dari metode *random forest*. Sehingga dapat dinyatakan bahwa metode *Weighted Random Forest* bekerja lebih baik dalam memprediksi pada kondisi *imbalanced data*. Hal ini didasarkan pada nilai *accuracy, precision, recall, F-Measure,* dan nilai AUC yang meningkat dari pada metode *random forest*.

Hasil Klasifikasi

Berdasarkan hasil analisis metode yang baik dalam klasifikasi adalah metode *Weighted Random Forest*. Dengan metode tersebut akan dilihat hasil prediksi untuk data *testing* dan data asli penelitian. Hasil prediksi data *testing* pada kelangsungan hidup pasien gagal jantung dengan menggunakan metode *weigthed random forest* terdapat 56 pasien gagal jantung diprediksi dengan benar dengan persentase sebesar 90,32% dan 6 pasien gagal jantung yang diprediksi salah dengan persentase sebesar 9,68%. Hasil prediksi klasifikasi kelangsungan hidup menggunakan metode *Weighted Random Forest* ada terdapat 228 pasien gagal jantung diprediksi dengan benar dengan persentase sebesar 93,061% dan 17 pasien gagal jantung diprediksi salah dengan persentase sebesar 6,939%. Selain itu dilakukan prediksi klasifikasi dengan menggunakan data baru dengan hasil prediksi yang disajikan pada Tabel 5.

Tabel 5 Prediksi Data Baru dengan Metode *Weighted Random Forest*

Umur	Jenis Kelamin	Hipertensi	Diabetes Melitus	Merokok	Prediksi
62	1	1	1	1	Selamat
38	0	1	1	1	Selamat

Umur	Jenis Kelamin	Hipertensi	Diabetes Melitus	Merokok	Prediksi
70	1	0	1	1	Selamat
67	0	0	1	1	Selamat
47	1	1	0	1	Selamat
35	0	1	0	1	Selamat
51	1	0	0	1	Selamat
70	0	0	0	1	Selamat
82	1	1	1	0	Meninggal
38	0	1	1	0	Selamat
65	1	0	1	0	Selamat
40	0	0	1	0	Selamat
58	1	1	0	0	Meninggal
39	0	1	0	0	Selamat
71	1	0	0	0	Meninggal
50	0	0	0	0	Selamat

Berdasarkan Tabel 5 menunjukkan terdapat 13 pasien gagal jantung yang diprediksi selamat dan 3 pasien gagal jantung yang diprediksi meninggal. Pasien yang diprediksi meninggal memiliki jenis kelamin yang sama yaitu pasien berjenis kelamin perempuan. 2 dari 3 pasien gagal jantung yang diprediksi meninggal berusia lansia > 60 tahun dan memiliki riwayat hipertensi. 1 dari 3 pasien yang diprediksi meninggal memiliki riwayat diabetes melitus. Sehingga dari data tersebut menunjukkan bahwa seorang pasien gagal jantung diprediksi meninggal dikarenakan usia yang sudah lansia, berjenis kelamin perempuan dan memiliki riwayat hipertensi.

D. Kesimpulan

Metode *Weighted Random Forest* lebih baik dibandingkan dengan metode *random forest* standar. Berdasarkan nilai kinerja klasifikasi metode *Weighted Random Forest* sudah baik dalam memprediksikan yang kelas pasien gagal jantung. Hasil evaluasi klasifikasi *Weighted Random Forest* untuk *accuracy* sebesar 90,32%, *precision* sebesar 93,22%, *recall* sebesar 96,49%, *F-Measure* sebesar 94,82%, dan nilai AUC sebesar 0,5825.

Daftar Pustaka

[1] Breiman, L. (2001). *Random Forests. Machine Learning*, 45:5-32. https://doi.org/10.1023/A:1010933404324

[2] Cutler, A., Cutler, D. R. & Stevens, J., (2011). *Random forests. Machine Learning*, 45(1), 157-176.

[3] Gorunescu, F. (2010). *Data Mining: Concepts, Models and Techniques*. Romania: Spinger. Retrieved Maret 15, 2022, from https://link.springer.com/book/10.1007/978-3-642-19721-5

[4] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Edition 3*. Burlington: Elsevier.

[5] Kementrian Kesehatan Republik Indonesia. (2019). *Hari Jantung Sedunia (World Heart Day): Your Heart is Our Heart Too*. Direktorat Pencegahan dan Pengendalian Penyakit Tidak Menular, (Online). (http://p2ptm.kemkes.go.id/kegiatan-p2ptm/pusat-/hari-jantung-sedunia-world-heart-day-your-heart-is-our-heart-too diakses pad tanggal 2 Maret 2022)

[6] Nugraha, I. S. (2017). *Karakteristik Pasien Gagal Jantung Rawat Inap Di RSUP Dr. Wahidin Sudirohusodo Makassar Periode 1 Januari – 30 September 2017*. Skripsi tidak dipublikasi. Makasar: Fakultas Kedokteran, Universitas Hasanuddin, Makasar.

[7] Rachman, F., & Purnami, S. W. (2012). *Perbandingan Klasifikasi Tingkat Keganasan*

Breast Cancer dengan Menggunakan Regresi Logistik Ordinal dan Support Vector Machine (SVM). Jurnal Sains dan Seni ITS. 1:130-135.

[8] Siringoringo, R. (2018). *Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor*. Journal Information System Development. 3:44-49.

[9] Tan, P.N., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining (First Edition)*. Pearson Addison Wesley, Boston.

[10] Timofeev, R. (2004). *Classification and Regression Trees (CART) Theory and Applications.* Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.

[11] Wong, T. T. & Yeh, S. J. (2019). *Weighted Random Forests for Evaluating Financial Credit Risk*. Proceedings of Engineering and Technology Innovation, 3:01-09.

[12] World Health Organization. (2020). *Cardiovascular diseases*. WHO World Health Organization, (Online), (https://www.who.int/health-topics/cardiovascular-diseases diakses pada tanggal 2 Maret 2022).

[13] Xu, H., Kong, Y. & Tan, S. (2020). *Predictive Modeling of Diabetic Kidney Disease using Random Forest Algorithm along with Features Selection*. Proceedings of the 2020 International Symposium on Artificial Intelligence in Medical Sciences.

[14] Zhu, M., Xia, J., Jin, X. & Yan, M. (2017). *Class Weights Random Forest Algorithm for Processing Class Imbalanced Medical Data*. IEEE Access, 6:4641-4652.

[15] Utami, Andi Nur Fadhilah, Suwanda. (2021). *Penggunaan Estimator Robust Reweighted Minimum Covariance Determinant pada Diagram Kontrol T2 Hotelling untuk Monitoring Penyebaran Covid-19 di Korea Selatan*, Jurnal Riset Statistika, 1(1), 63-72.