

Evaluasi Performa Algoritma Klasifikasi dalam Prediksi Kekambuhan Kanker Tiroid Pasca Terapi RAI: Studi Kasus *Dataset RAI Therapy*

Wahyu Nugraha¹, Raja Sabaruddin²

^{1,2}Universitas Bina Sarana Informatika, Indonesia
Email: wahyu.whn@bsi.ac.id¹, raja.rjd@bsi.ac.id²

Alamat: Jl. Abdul Rahman Saleh No.18, Bangka Belitung Laut, Kec. Pontianak Tenggara, Kota Pontianak, Kalimantan Barat 78124

Korespondensi Penulis: wahyu.whn@bsi.ac.id

Abstract. *Thyroid cancer is the most common endocrine malignancy, with a steadily increasing incidence rate. Although the overall survival rate is relatively high, the risk of recurrence after definitive treatment such as Radioactive Iodine (RAI) therapy remains a significant clinical challenge. Predicting recurrence risk is crucial for optimizing monitoring strategies and interventions. With advances in technology, machine learning (ML) approaches are increasingly utilized to support medical predictions, including the recurrence of thyroid cancer. This study aims to evaluate the performance of four classification algorithms—Logistic Regression, XGBClassifier, Random Forest Classifier, and Voting Classifier—in predicting thyroid cancer recurrence using the Thyroid Cancer Recurrence After RAI Therapy dataset, which consists of 383 patient records and 13 key clinical attributes. The evaluation was conducted using accuracy, precision, recall, F1-score, and area under the curve (AUC) metrics. The results show that the XGBClassifier is the best-performing model with an accuracy of 97.4% and an AUC of 0.95, demonstrating superior performance in handling the minority class. This research is expected to contribute to the development of more effective machine learning-based clinical decision support systems for predicting thyroid cancer recurrence after therapy.*

Keywords: *Thyroid Cancer, Recurrence, Machine Learning, Classification*

Abstrak. Kanker tiroid merupakan kanker endokrin yang paling umum dengan angka kejadian yang terus meningkat. Meskipun tingkat kesembuhan relatif tinggi, risiko kekambuhan pasca terapi definitif seperti *Radioactive Iodine (RAI) therapy* tetap menjadi tantangan klinis yang serius. Prediksi risiko kekambuhan menjadi krusial untuk mengoptimalkan strategi pemantauan dan intervensi. Seiring dengan perkembangan teknologi, pendekatan *machine learning (ML)* semakin banyak digunakan untuk mendukung prediksi medis, termasuk dalam memprediksi kekambuhan kanker tiroid. Penelitian ini bertujuan untuk mengevaluasi performa empat algoritma klasifikasi yaitu *Logistic Regression, XGBClassifier, Random Forest Classifier, dan Voting Classifier* dalam memprediksi kekambuhan kanker tiroid menggunakan dataset *Thyroid Cancer Recurrence After (RAI) Therapy* yang terdiri dari 383 data pasien dengan 13 atribut klinis utama. Evaluasi dilakukan berdasarkan metrik akurasi, *precision, recall, F1-score*, dan *area under the curve (AUC)*. Hasil penelitian menunjukkan bahwa *XGBClassifier* merupakan model terbaik dengan akurasi sebesar 97,4% dan AUC sebesar 0,95, serta memiliki keunggulan dalam menangani kelas minoritas. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem pendukung keputusan klinis berbasis *machine learning* yang lebih efektif dalam memprediksi kekambuhan pasien kanker tiroid pasca terapi.

Kata kunci: Kanker tiroid, *Recurrence, Machine Learning, Algoritma Klasifikasi*

1. LATAR BELAKANG

Kanker tiroid merupakan jenis kanker endokrin yang paling umum, dengan angka kejadian yang terus meningkat dalam beberapa dekade terakhir (Siegel et al., 2020). Meskipun tingkat kesembuhan pasien kanker tiroid tergolong tinggi, terdapat sebagian pasien tetap berisiko mengalami kekambuhan (*recurrence*) setelah menjalani terapi definitif seperti *Radioactive Iodine (RAI) therapy*. Kekambuhan ini menjadi tantangan

klinis yang serius karena dapat berdampak pada kualitas hidup dan prognosis jangka panjang.

Prediksi risiko kekambuhan pasca terapi menjadi aspek penting dalam menentukan strategi pemantauan dan intervensi lebih lanjut. Seiring berkembangnya teknologi, pendekatan berbasis *machine learning* (ML) telah banyak digunakan untuk mendukung prediksi medis, termasuk dalam memprediksi kekambuhan kanker tiroid (Deo, 2015). Berbagai penelitian sebelumnya telah menerapkan algoritma klasifikasi seperti *Support Vector Machine* (SVM), *k-Nearest Neighbors* (k-NN), *Decision Tree*, hingga *Neural Networks*, untuk mengolah data klinis dan menghasilkan model prediksi kekambuhan (Obaido et al., 2024), (Iqbal & Shahzad, 2024). Namun, kinerja masing-masing algoritma dapat sangat bergantung pada karakteristik data, teknik pra-pemrosesan, dan parameter model (Mohammed et al., 2025). Misalnya, pemilihan teknik penskalaan data yang tepat dapat secara signifikan mempengaruhi performa model klasifikasi, dan penggunaan teknik yang tidak sesuai bahkan dapat lebih merugikan daripada tidak melakukan penskalaan sama sekali. Selain itu, kualitas data seperti kelengkapan dan konsistensi juga memainkan peran penting dalam menentukan efektivitas model *machine learning*. Lebih lanjut, proses *tuning hyperparameter*, baik pada tahap pra-pemrosesan maupun pada algoritma itu sendiri, merupakan faktor krusial yang dapat mempengaruhi hasil akhir model (de Amorim et al., 2023). Oleh karena itu, evaluasi komprehensif terhadap kinerja berbagai algoritma klasifikasi pada *dataset* spesifik menjadi penting sebelum implementasi lebih lanjut dalam sistem pendukung keputusan klinis.

Dalam penelitian ini, dilakukan evaluasi performa empat algoritma klasifikasi, yaitu *Logistic Regression*, *XGBClassifier*, *Random Forest Classifier*, dan *Voting Classifier* dalam memprediksi kekambuhan kanker tiroid menggunakan *dataset Thyroid Cancer Recurrence After Radioactive Iodine (RAI) Therapy*. *Logistic Regression* dipilih karena kesederhanaannya serta interpretabilitas model yang tinggi (Hosmer et al., 2013), sedangkan *XGBClassifier* merupakan algoritma *boosting* modern yang dikenal memiliki akurasi tinggi pada data tabular (Chen & Guestrin, 2016). *Random Forest Classifier* digunakan karena kemampuannya dalam mengurangi *overfitting* melalui penggabungan banyak *decision tree* (Biau & Scornet, 2016), sementara *Voting Classifier* digunakan untuk mengombinasikan kekuatan beberapa model sekaligus guna meningkatkan performa prediksi (Rokach, 2010). Dari beberapa model yang diuji langkah selanjutnya yaitu mengidentifikasi model terbaik yang tidak hanya akurat tetapi juga stabil untuk digunakan dalam skenario prediksi kekambuhan kanker tiroid. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan klinis berbasis *machine learning*, yang lebih handal dalam memprediksi kekambuhan pasien kanker tiroid pasca terapi RAI.

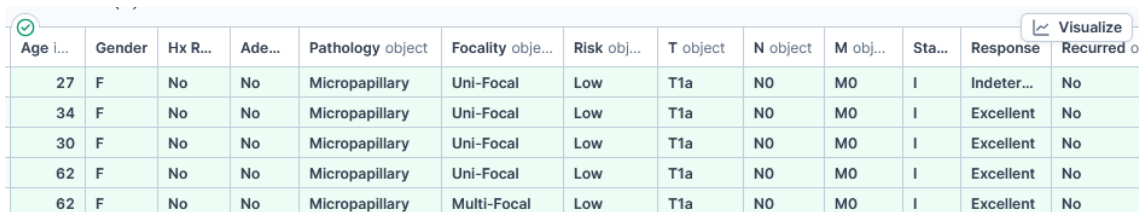
Penelitian ini menggunakan kumpulan data atau *dataset Thyroid cancer recurrence after Radioactive Iodine (RAI) therapy*. *Dataset* ini mencakup 383 rekaman pasien dengan 13 atribut utama, termasuk usia, jenis kelamin, stadium kanker, tipe patologi, klasifikasi risiko, respons pengobatan, dan status kekambuhan. Data ini berguna untuk memprediksi kekambuhan kanker, memahami faktor risiko, dan mengevaluasi hasil pengobatan. *Dataset* yang digunakan merupakan *dataset public* yang diperoleh dari <https://www.kaggle.com/datasets/aneevinay/thyroid-cancer-recurrence-dataset>. Evaluasi mencakup pengukuran metrik performa seperti akurasi, *precision*, *recall*, *F1-score*, dan

area under the curve (AUC) untuk memberikan gambaran menyeluruh mengenai kekuatan dan keterbatasan masing-masing algoritma. Hasil eksperimen menunjukkan bahwa model Model *XGBClassifier* direkomendasikan sebagai model terbaik karena memperoleh akurasi tertinggi yaitu sebesar 97.4% dan AUC sebesar 0,95. Model *XGBClassifier* terbaik secara keseluruhan, terutama karena mampu menangani kelas minoritas dengan baik.

2. KAJIAN TEORITIS

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terkait dengan *Thyroid cancer recurrence after Radioactive Iodine (RAI) therapy* yang diperoleh dari *Kaggle dataset repository* dalam format CSV (*comma separated values*). *Dataset* ini berfokus pada kekambuhan kanker tiroid setelah terapi *Iodin Radioaktif (RAI)*. *Dataset* ini berisi 383 catatan pasien dengan 13 atribut diantaranya yaitu *age, gender, cancer staging, pathology type, risk classification, treatment response, dan recurrence status*. Data ini sangat berguna untuk memprediksi kekambuhan kanker, memahami faktor risiko, serta mengevaluasi hasil pengobatan. Gambar 1 merupakan informasi sederhana dari *dataset*. Menggunakan metode untuk mencetak informasi tentang sebuah *DataFrame*, termasuk tipe data dari indeks dan kolom.



Age	Gender	Hx R...	Ade...	Pathology object	Focality obje...	Risk obj...	T object	N object	M obj...	Sta...	Response	Recurred o...
27	F	No	No	Micropapillary	Uni-Focal	Low	T1a	N0	M0	I	Indeter...	No
34	F	No	No	Micropapillary	Uni-Focal	Low	T1a	N0	M0	I	Excellent	No
30	F	No	No	Micropapillary	Uni-Focal	Low	T1a	N0	M0	I	Excellent	No
62	F	No	No	Micropapillary	Uni-Focal	Low	T1a	N0	M0	I	Excellent	No
62	F	No	No	Micropapillary	Multi-Focal	Low	T1a	N0	M0	I	Excellent	No

Gambar 1. Informasi *Dataframe*

Validasi

10-fold Cross-Validation merupakan metode standar dalam evaluasi model *machine learning* untuk menilai kinerja model dengan lebih presisi sekaligus mengurangi risiko *overfitting*. Pada teknik ini, data dibagi menjadi 10 bagian (*fold*) yang sama besar. Secara bergiliran, model dilatih menggunakan 9 *fold* dan diuji pada 1 *fold* yang tersisa. Proses ini diulang hingga semua *fold* pernah menjadi data uji. Hasil akhirnya dirata-rata dari 10 kali pengujian, menghasilkan estimasi performa model yang lebih konsisten dibandingkan evaluasi satu kali pemisahan data (Normawati & Ismi, 2019). Metode ini dikenal efektif karena memastikan bahwa setiap data digunakan baik untuk pelatihan maupun pengujian, sehingga meminimalkan variabilitas hasil yang muncul akibat pembagian dataset yang berbeda. Teknik ini terutama bermanfaat pada *dataset* yang relatif kecil, karena memastikan semua data berkontribusi dalam proses pelatihan dan pengujian model (Nti et al., 2021). Tabel 1 menggambarkan proses *10-fold cross-validation* dimana setiap iterasi menggunakan satu *fold* sebagai *test fold*, sementara *fold* lainnya digunakan sebagai *training folds*.

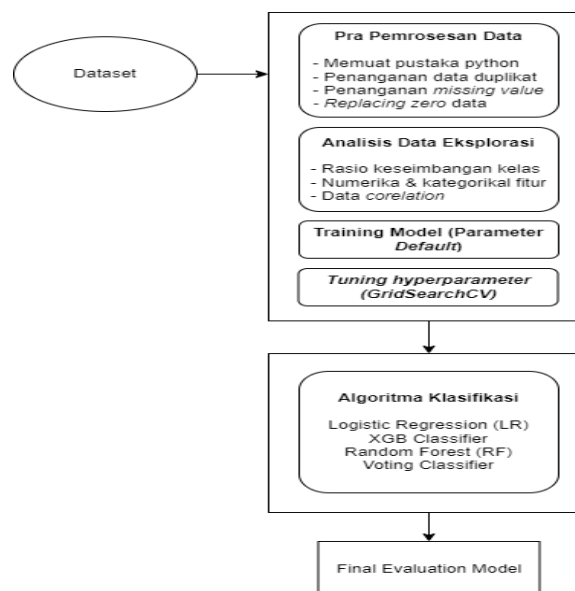
Tabel 1. Proses *10-fold cross-validation*

Iteration	Training Folds	Test Fold
Fold 1	Fold 2-10	Fold 1
Fold 2	Fold 1,3-10	Fold 2

Fold 3	Fold 1-2,4-10	Fold 3
Fold 4	Fold 1-3,5-10	Fold 4
Fold 5	Fold 1-4,6-10	Fold 5
Fold 6	Fold 1-5,7-10	Fold 6
Fold 7	Fold 1-6,8-10	Fold 7
Fold 8	Fold 1-7,9-10	Fold 8
Fold 9	Fold 1-8,10	Fold 9
Fold 10	Fold 1-9	Fold 10

3. METODE PENELITIAN

Pada penelitian ini eksperimen model *machine learning* menggunakan beberapa algoritma klasifikasi. Gambar 2 Menunjukkan tahapan eksperimen yang dilakukan mulai dari *import* dataset menggunakan *library* python, data *preprocessing*, *exploratory data analysis* (EDA), uji coba model menggunakan beberapa algoritma klasifikasi dan tahap akhir akan dilakukan evaluasi performa model.



Gambar 2. Alur Tahapan Penelitian

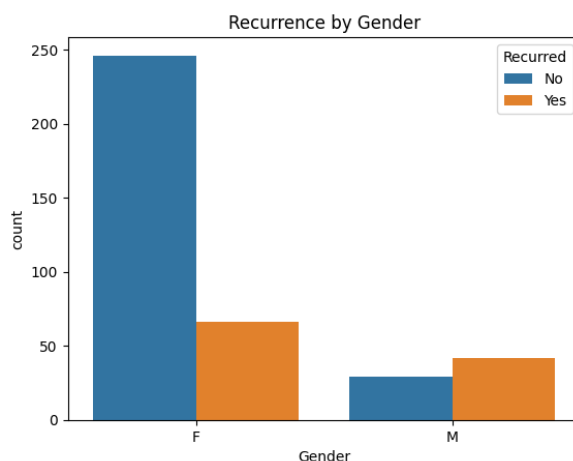
4. HASIL DAN PEMBAHASAN

Eksperimen terhadap model yang dikembangkan dilakukan menggunakan bahasa pemrograman *Python* versi 3.9 melalui lingkungan pengembangan *Deeptime*. Berbagai pustaka atau *library* digunakan dalam proses ini, antara lain *sklearn*, *numpy*, *pandas*, *matplotlib*, *seaborn*, *plotly*, dan beberapa lainnya. Proses pra-pemrosesan data (*preprocessing*) dilakukan dengan bantuan *library* *pandas*, dimulai dengan memuat dataset berformat CSV. Setelah dataset dimuat, dilakukan observasi awal untuk mendeteksi adanya duplikasi data serta memeriksa apakah terdapat data yang hilang (*missing value*).

Exploratory Data Analysis

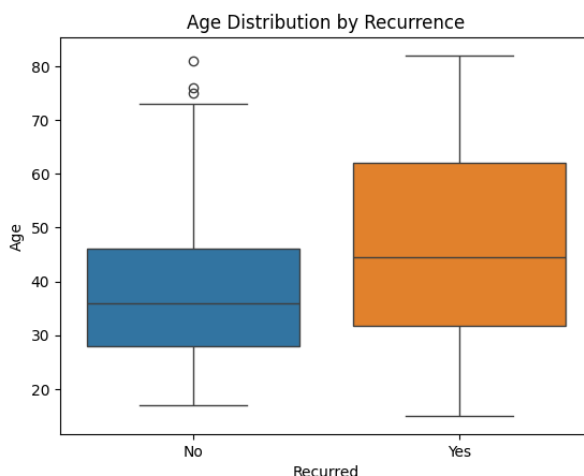
Langkah pertama dilakukan visualisasi data untuk menggali informasi terhadap karakteristik *dataset* yang digunakan. Gambar 3 menggambarkan data kekambuhan

kanker (*recurrence*) berdasarkan jenis kelamin. Pada gambar diagram batang tersebut menunjukkan jumlah kekambuhan kanker tiroid berdasarkan jenis kelamin pasien setelah menjalani terapi *Iodin Radioaktif*. Dari gambar tersebut dapat diambil kesimpulan bahwa perempuan lebih banyak jumlahnya dalam *dataset*, namun lebih banyak yang tidak mengalami kekambuhan. Sedangkan Laki-laki memiliki kecenderungan kekambuhan yang lebih tinggi secara proporsional, meskipun jumlah totalnya lebih sedikit.



Gambar 3. Visualisasi kekambuhan kanker berdasarkan gender

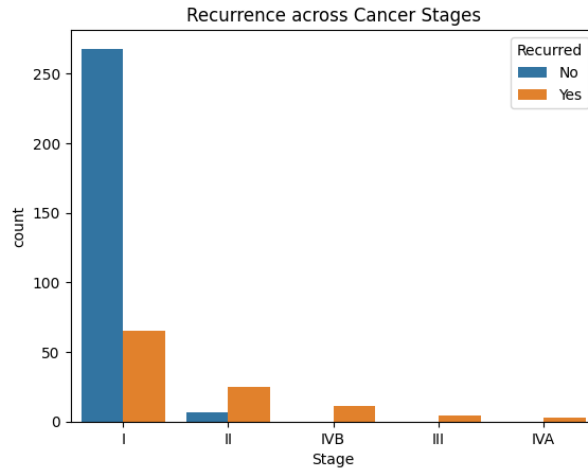
Gambar 4 merupakan *boxplot* yang menggambarkan distribusi usia pasien berdasarkan status kekambuhan kanker tiroid (*recurred*) setelah terapi *Iodin Radioaktif*. Dari gambar tersebut dapat disimpulkan bahwa Pasien yang mengalami kekambuhan cenderung memiliki usia lebih tua dibanding pasien yang tidak kambuh. Kemudian distribusi usia pasien yang kambuh juga lebih menyebar luas, menunjukkan bahwa kekambuhan bisa terjadi di berbagai rentang usia meskipun mediannya lebih tinggi. Grafik Ini dapat menunjukkan bahwa usia mungkin merupakan faktor risiko kekambuhan kanker tiroid.



Gambar 4. Boxplot distribusi berdasarkan usia

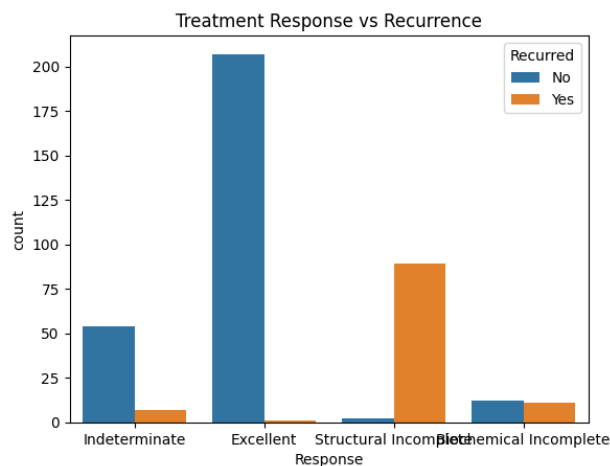
Gambar 5 merupakan diagram batang yang menunjukkan jumlah kekambuhan kanker tiroid berdasarkan stadium kanker (*stage*). Dari gambar tersebut dapat dijelaskan bahwa mayoritas pasien berada di stadium I, dan sebagian besar dari mereka tidak

mengalami kekambuhan. Proporsi kekambuhan meningkat pada stadium yang lebih tinggi, terutama mulai dari stadium II ke atas. Stadium kanker tampaknya berkorelasi positif dengan risiko kekambuhan: makin lanjut stadium, makin tinggi kemungkinan kekambuhan.



Gambar 5. kekambuhan kanker tiroid berdasarkan stadium kanker (*stage*)

Gambar 6 merupakan diagram batang yang menunjukkan hubungan antara respons terhadap pengobatan dengan status kekambuhan (*recurrence*) kanker tiroid setelah terapi *Radioactive Iodine*. Dari gambar tersebut menunjukkan bahwa respons pengobatan sangat berpengaruh terhadap kekambuhan khususnya pasien dengan respons *Excellent* yang hampir tidak mengalami kekambuhan jika dibandingkan dengan pasien dengan respons *Incomplete (Structural/Biochemical)* berisiko tinggi mengalami kekambuhan. Ini memperkuat pentingnya evaluasi respons pengobatan untuk memprediksi risiko kekambuhan kanker tiroid.

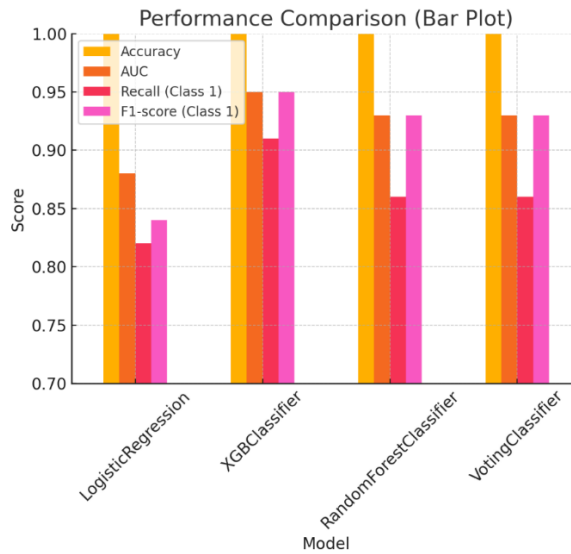


Gambar 6. Hubungan antara respons pengobatan dengan status (*recurrence*)

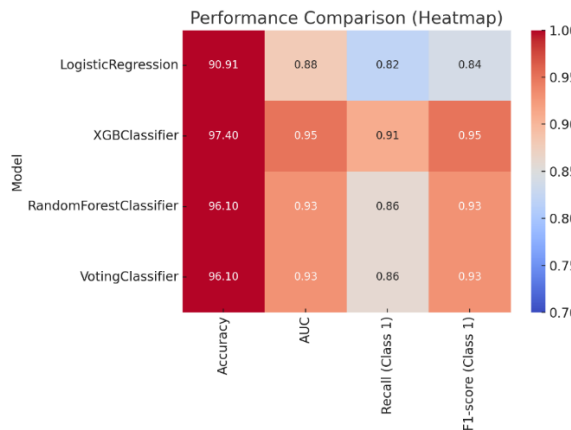
Hasil Penelitian

Hasil dari pengujian model dapat di lihat pada gambar 7 dan 8. Gambar 7 merupakan diagram batang (*bar plot*) yang menunjukkan perbandingan performa empat model klasifikasi berdasarkan empat metrik evaluasi. Nilai *accuracy* terlihat hampir sempurna (mendekati 1.00) untuk semua model. Artinya, model memiliki akurasi tinggi

dalam klasifikasi data keseluruhan. Meskipun keempat model menunjukkan akurasi tinggi, *XGBClassifier* unggul dalam mendeteksi kasus kekambuhan (*Class 1*), sebagaimana ditunjukkan oleh *recall* dan *F1-score* yang tinggi. Gambar 8 merupakan heatmap perbandingan performa model klasifikasi berdasarkan beberapa metrik evaluasi. Dari kedua gambar tersebut dapat disimpulkan bahwa *XGBClassifier* adalah model terbaik, karena tidak hanya mencapai akurasi dan AUC sempurna, tetapi juga memiliki performa terbaik dalam mendeteksi pasien dengan kekambuhan (*Class 1*), yang sangat krusial dalam konteks prediksi medis. Model lainnya seperti *Random Forest* dan *Voting Classifier* tetap memberikan performa yang baik dan seimbang sehingga bisa menjadi alternatif yang sangat baik, sedangkan *Logistic Regression* kurang cocok jika fokus utamanya adalah mengidentifikasi kekambuhan secara akurat.



Gambar 7. Bar plot perbandingan performa model klasifikasi



Gambar 8. Heatmap perbandingan performa model klasifikasi

5. KESIMPULAN DAN SARAN

Kebutuhan akan prediksi dini kekambuhan kanker tiroid sangat penting untuk mendukung pengambilan keputusan klinis pasca terapi RAI, mengingat kekambuhan dapat mempengaruhi kualitas hidup dan prognosis pasien. Pendekatan *machine learning* efektif dalam memprediksi kekambuhan, dengan berbagai algoritma memiliki performa

yang bervariasi tergantung pada teknik pra-pemrosesan dan karakteristik data. *XGBClassifier* terbukti sebagai model paling unggul dalam penelitian ini, dengan akurasi sebesar 97,4% dan AUC sebesar 0,95. Model ini juga menunjukkan performa yang baik dalam menangani kelas minoritas, sehingga direkomendasikan untuk digunakan dalam sistem pendukung keputusan klinis.

Pada penelitian selanjutnya akan jauh lebih baik jika validasi eksternal menggunakan *dataset* dari institusi medis lain untuk menguji generalisasi model dalam populasi yang berbeda. Kemudian pertimbangkan penggunaan teknik *balancing* data seperti *SMOTE* atau *ADASYN* untuk mengatasi potensi ketidakseimbangan kelas secara lebih eksplisit. Terapkan *Explainable AI* (XAI) untuk meningkatkan transparansi model dan memahami kontribusi setiap fitur terhadap prediksi, sehingga lebih mudah diterima dalam praktik klinis. Untuk implementasi dari model yang dibuat kembangkan antarmuka klinis berbasis *web* atau aplikasi, yang memungkinkan dokter untuk mengakses dan menggunakan model prediksi ini secara *real-time* dengan input data pasien.

DAFTAR REFERENSI

- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- de Amorim, L. B. V, Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing*, 133, 109924. <https://doi.org/https://doi.org/10.1016/j.asoc.2022.109924>
- Deo, R. C. (2015). Machine Learning in Medicine. *Circulation*, 132(20), 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.
- Iqbal, D., & Shahzad, A. (2024). Prediction of thyroid cancer recurrence with machine learning models. *Pakistan Journal of Nuclear Medicine*, 14(2), 49–55. <https://doi.org/10.24911/PJNMed.175-1721068107>
- Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems*, 132, 102549. <https://doi.org/https://doi.org/10.1016/j.is.2025.102549>
- Normawati, D., & Ismi, D. P. (2019). K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining. *Signal and Image Processing Letters*, 1(2), 23–35. <https://doi.org/10.31763/simple.v1i2.3>
- Nti, I., Nyarko-Boateng, O., & Aning, J. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation. *International Journal of Information Technology and Computer Science*, 6, 61–71. <https://doi.org/10.5815/ijitcs.2021.06.05>
- Obaido, G., Achilonu, O., Ogbuokiri, B., Amadi, C. S., Habeebullahi, L., Ohalloran, T., Chukwu, C. W., Mienye, E. D., Aliyu, M., Fasawe, O., Modupe, I. A., Omietimi, E. J., & Aruleba, K. (2024). An Improved Framework for Detecting Thyroid Disease

- Using Filter-Based Feature Selection and Stacking Ensemble. *IEEE Access*, 12, 89098–89112. <https://doi.org/10.1109/ACCESS.2024.3418974>
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1), 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
- Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7–30. <https://doi.org/10.3322/caac.21590>