MAPLE
Mechatronics Journal in Professional and Entrepreneur
Volume 7 Edisi 2 Bulan Desember Tahun 2025 Politeknik Bosowa P-ISSN : 2685-4868, E-ISSN : 2963-6175

Pengaruh *Base Estimator* Pada Algoritma *Adaptive Boost* Pada Prediksi Penyakit Akut

Fahrim Irhamna Rahman[#], Muhyiddin A.M Hayat, Muh. Rasdi
Program Studi Informatika, Universitas Muhammadiyah Makassar
Jl. Sultan Alauddin No.259, Gn. Sari, Kec. Rappocini, Kota Makassar, Sulawesi Selatan 90221, Indonesia
rasdyardia@gmail.com

Abstrak

Penyakit akut merupakan kondisi medis yang muncul secara tiba-tiba dan memerlukan penanganan cepat untuk mencegah komplikasi serius. Prediksi penyakit akut yang akurat menjadi tantangan dalam dunia medis, terutama dengan meningkatnya jumlah data pasien yang perlu dianalisis secara efisien. Penelitian ini bertujuan untuk mengevaluasi pengaruh berbagai jenis base estimator pada algoritma Adaptive Boost dalam meningkatkan akurasi prediksi penyakit akut. Metode yang digunakan dalam penelitian ini adalah Adaptive Boost dengan tiga jenis base estimator, yaitu Decision Tree, Support Vector Classifier (SVC), dan Logistic Regression. Data yang digunakan berasal dari rekam medis pasien di Rumah Sakit Umum Daerah Mamuju, dengan jumlah sampel sebanyak 1004 data pasien. Data mengalami proses preprocessing, termasuk cleaning, transformation, dan feature selection. Model yang dibangun kemudian dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Logistic Regression sebagai base estimator menghasilkan akurasi tertinggi sebesar 93%, diikuti oleh Decision Tree dan SVC, masing-masing dengan akurasi 89%.

Kata kunci: *Adaptive Boost, Base Estimator,* Prediksi Penyakit Akut, *Machine Learning, Ensemble Learning.*

Abstract

Acute diseases are medical conditions that arise suddenly and require prompt treatment to prevent severe complications. Accurate prediction of acute diseases remains a challenge in the medical field, especially with the increasing amount of patient data that needs to be analyzed efficiently. This study aims to evaluate the effect of different types of base estimators on the Adaptive Boost algorithm in improving the accuracy of acute disease prediction. The method used in this research is Adaptive Boost with three types of base estimators: Decision Tree, Support Vector Classifier (SVC), and Logistic Regression. The dataset consists of 1004 patient records from the Mamuju Regional General Hospital. The data underwent preprocessing steps, including cleaning, transformation, and feature selection. The developed models were then evaluated using accuracy, precision, recall, and F1-score metrics. The results show that the Logistic Regression model as a base estimator achieved the highest accuracy of 93%, followed by Decision Tree and SVC, both with an accuracy of 89%..

Keywords: *Adaptive Boost, Base Estimator, Acute Disease Prediction, Machine Learning, Ensemble Learning.*

I. PENDAHULUAN

Penyakit akut adalah kondisi kesehatan yang muncul secara tiba-tiba, berlangsung dalam waktu yang relatif singkat, dan sering kali ditandai dengan gejala yang parah atau intens. Berbeda dengan penyakit kronis yang berkembang secara perlahan dan bertahan lama, penyakit akut biasanya memiliki durasi yang singkat dan memerlukan penanganan segera. Penyakit akut dapat disebabkan oleh banyak hal, seperti infeksi, cedera, dan kondisi medis lainnya. Jika Anda mengalami gejala yang tiba-tiba dan parah, sangat penting untuk mendapatkan perawatan medis segera karena penyakit akut biasanya memerlukan perawatan cepat dan dapat berakibat fatal jika tidak diobati tepat waktu.

Penelitian ini bertujuan untuk memberikan kontribusi dengan menggunakan algoritma pembelajaran mendalam dalam memprediksi penyakit akut seperti demam berdarah (DBD), tifus, Gastroenteritis, infeksi saluran pernapasan atas (ISPA), dan infeksi saluran kemih (ISK). Pemilihan penyakit tersebut didasarkan pada tingginya angka dan tantangan dalam diagnosis, yang seringkali memerlukan identifikasi yang cepat dan tepat untuk mencegah komplikasi serius.

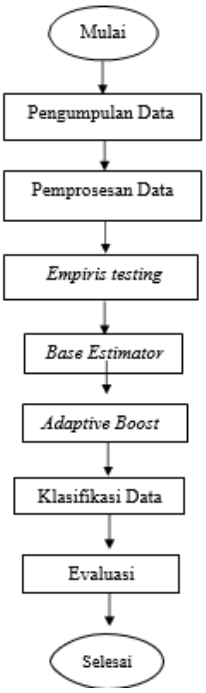
Meskipun metode umum dalam analisis data medis telah memberikan sumbangan yang berharga, tantangan untuk meningkatkan akurasi dan kehandalan prediksi penyakit akut masih ada. Oleh karena itu, pengunaan metode Ensemble Learning menjadi relevan. Metode ini merupakan pendekatan dalam machine learning yang menggabungkan beberapa model untuk meningkatkan kinerja. Salah satu teknik yang sering digunakan adalah Adaptive Boost yang merupakan teknik boosting dengan cara menggunakan metode secara besama-sama dalam pembelajaran mesin untuk meningkatkan akurasi [1].

Pada algoritma *Adaptive Boost*, *ensemble* learning merujuk pada gabungan beberapa weak learner atau classifier yang kurang efektif dengan satu learner yang memiliki kemampuan prediksi yang baik. Adaptive Boost memiliki keunggulan dalam menggabungkan berbagai macam model algoritma, seperti Naïve Bayes, Decision Tree, Logistic Regression, dan sebagainya [2]. Dengan demikian, diharapkan penggunaan metode ini mampu menghasilkan model yang lebih dapat diandalkan dan dapat dipercaya dalam memprediksi penyakit akut, mempermudah tindakan yang tepat waktu dan efektif untuk meningkatkan perkiraaan keseluruhan kondisi pasien.

Dalam konteks pembelajaran mesin (machine learning), Base Estimator merujuk pada model dasar yang digunakan sebagai komponen pembangun dalam algoritma ensemble. Algoritma *ensemble* menggabungkan beberapa model atau estimator untuk mencapai performa yang lebih baik dibandingkan dengan satu model tunggal. Penggabungan beberapa base estimator dapat mengurangi kesalahan prediksi dan meningkatkan akurasi secara keseluruhan.

II. METODE PENELITIAN

Metode penelitian ini melibatkan beberapa tahapan yang disusun secara sistematis untuk mencapai tujuan penelitian,. Tahap penelitian ini dapat di lihat dari f*lowchart* yang membantu dalam mengorganisir informasi, memvisualisasikan hubungan antar proses, dan menyediakan panduan yang dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Penelitian

Sesuai dengan gambar diatas, tahapan penelitian terbagi menjadi beberapa tahapan yaitu sebagai berikut :

a. Pengumpulan data

Pengumpulan data dilakukan dengan mengambil dataset dari Rumah Sakit Umum Daerah (RSUD) Sulawesi Barat. yang berisi informasi rekam medis pasien penyakit. Data yang diambil dari ruang rekam medis ialah data mentah biasanya berupa data rekam medis yang dikumpulkan dari rumah sakit, laboratorium, atau sumber lain berupa file teks (Excel), data rekam medis elektronik (EHR) dan observasi manual. Informasi yang diperoleh meliputi berbagai kategori, yaitu nomor rekam medis, nama penyakit serta,

gejala yang dialami pasien. Setelah data terkumpul seluruhnya di masukkan kedalam data excel.

NO	RM	PENYAKIT	GEJALA
1	62699	THYPOID	LEMAS, DEMAM (-), KDG SAKIT KEPALA
2	18426	ISPA	BATUK (+), PILEK (+), DEMAM (-), RIWAYAT DEMAM (+), NYERI ULU HATI (+), ,MALAS MAKAN (+), , NYERI TELINGA (+) KIRI
3	67537	ISPA	BAtuk berdahak (+) 3hari, demam kurang lebih 2 hari
4	27804	GASTROENTERITIS AKUT	bab encer berkurang,nyeri ulu hati
5	99126	ISPA	batuk sekali2
6	100169	ISPA	batuk berdahak
7	99943	ISPA	batuk berdahak
8	96915	GASTROENTERITIS AKUT	BAB encer 2x, warnakuning kecoklatan
9	83159	ISPA	BATUK (+) DEMAM (+), PILEK (+), MALAS MAKAN (+)

Gambar 2. Data Awal

b. *Preprocessin* Data

Preprocessing data adalah tahap penting dalam analisis data yang bertujuan untuk mempersiapkan dataset agar siap untuk digunakan dalam proses analisis atau pengembangan model. Berikut adalah tahap mengenai setiap langkah dalam *preprocessing* data:

1. Cleaning

Tahap *Cleaning* data melibatkan proses menghapus tanda baca seperti koma, titik, tanda tanya, tanda seru, bintang, dan pagar dari teks atau data. Langkah ini penting karena karakter-karakter tersebut biasanya tidak memberikan kontribusi signifikan dalam analisis data atau pemrosesan teks. Tanda baca sering dianggap sebagai noise yang tidak relevan dan dihapus untuk memfokuskan perhatian pada informasi utama dalam teks tersebut.

2. Transformasi Data

Transformasi data adalah proses mengubah atau memodifikasi data mentah menjadi format yang lebih terstruktur dan sesuai untuk analisis atau pemodelan Pada tahap ini dilakukan transformasi data agar lebih mudah diproses oleh model. Untuk analisis gejala penyakit, gunakan kolom seperti Demam, Durasi Demam, Keparahan Demam, dan sebagainya. Kolom seperti Nama Penyakit digunakan untuk klasifikasi atau analisis label. Abaikan kolom yang tidak relevan seperti NO atau RM, kecuali dibutuhkan untuk identifikasi. Untuk kolom dengan nilai numerik, seperti Demam, Keparahan Demam atau Sakit Perut, Memetakan nilai ke dalam rentang (0 atau 1). sedangkan keparahan gejala direpresentasikan dalam kategori Ringan dan Parah.

3. Encoding Label

Mengonversi label target y yang berbentuk kategori (misalnya "sehat", "sakit") menjadi representasi numerik (misalnya 0, 1). LabelEncoder adalah bagian dari sklearn.preprocessing. fit_transform(y) memetakan setiap kategori unik ke nilai numerik (integer) dan menghasilkan label yang telah dikodekan. Contohnya Input: y = ["sehat", "sakit", "sehat", "sakit"], Output: y_encoded = [1, 0, 1, 0].

4. Normalisasi

Menyelaraskan nilai fitur x ke rentang 0-1 untuk memastikan semua fitur memiliki skala yang sama. MinMaxScaler adalah bagian dari sklearn.preprocessing. Semua nilai dalam kolom data x akan dipetakan ke rentang 0 hingga 1. Dalam proses ini memili manfaaat untuk mengurangi fitur dengan skala besar.

5. Split Data

Tujuannya ialah membagi data menjadi data latih (training set) dan data uji (test set). Sebagai parameter, test_size=0.2: 20% data akan digunakan sebagai data uji, sedangkan 80% untuk pelatihan. Kemudian, random_state=42: Menetapkan nilai acak tetap sehingga pembagian data selalu konsisten. Outputnya yaitu x_train, y_train ialah Data untuk melatih model dan x_test, y_test ialah Data untuk menguji performa model.

6. SMOTE

SMOTE bertujuan untuk menangani ketidakseimbangan kelas pada data latih. SMOTE membuat data sintetis untuk kelas minoritas (kelas dengan jumlah sampel lebih sedikit) menggunakan interpolasi. Dengan membuat data baru, distribusi kelas menjadi lebih seimbang. SMOTE juga mengurangi bias model terhadap kelas mayoritas serta mseningkatkan performa prediksi pada kelas minoritas. Contoh sebelum SMOTE y_train: [0, 0, 0, 1] (kelas mayoritas = 0) dan setelah SMOTE y_train: [0, 0, 0, 1, 1, 1] (jumlah kelas 0 dan 1 menjadi seimbang).

NO	RM	Nama Penyakit	Demam	Durasi Demam	Keparahan Demam	Sakit Tenggorokan	Durasi sakit	...
1	059434	Thypoid	1	4 hari	Parah	0	0	...
2	053815	DBD	1	0	Ringan	0	0	...
3	103750	Gastroenteritis	0	0	N/A	0	0	...
4	104297	ISPA	0	0	N/A	0	0	...
...	...	...	...	...	...	...	...	...
1001	054993	Gastroenteritis	0	0	0	0	0	...
1002	082484	ISK	0	0	0	0	0	...
1003	080580	ISPA	1	4 hari	2	0	0	...
1004	093101	ISPA	0	0	0	0	0	...

Gambar 3. Hasil Transformasi Data

c. *Empiris Testing*

Empiris testing adalah pendekatan berbasis bukti untuk mempelajari dan menafsirkan informasi. Bukti empiris adalah informasi yang dapat dikumpulkan dari pengalaman atau melalui panca indera.

d. Base Estimator

Base estimator adalah model atau algoritma dasar yang digunakan sebagai komponen utama dalam metode pembelajaran ensemble, seperti bagging, boosting, atau stacking. Base Estimator merupakan model pembelajaran mesin individu yang nantinya digabungkan untuk membentuk model yang lebih kompleks dan kuat. Penggunaan Base Estimator pada algoritma AdaBoost. Dalam Adaptive Boost, Base Estimator umumnya merupakan model yang lebih sederhana, seperti decision stump (pohon keputusan sederhana dengan satu split). Base Estimator adalah komponen fundamental yang dilatih dan kemudian digabungkan dalam teknik ensemble untuk meningkatkan kinerja model secara keseluruhan. Ketika kita membangun model ensemble, kita dapat menentukan jenis Base Estimator yang digunakan dengan memilih algoritma sederhana yaitu Dececion Tree, SVC, dan logistic Regretion.

e. Adaptive Boost

Adaptive Boost (AdaBoost) adalah salah satu variasi dari algoritma boosting. Menurut penelitian oleh Freund dan Schapire pada tahun 1999, algoritma boosting bertujuan untuk mengubah model lemah (weak learner) menjadi model yang lebih kuat (strong learner). Secara umum, boosting berfokus pada pembentukan serangkaian pohon keputusan menggunakan base learner tertentu. Inti dari algoritma AdaBoost adalah memberikan bobot yang lebih besar pada observasi yang salah klasifikasi [3].

Menurut Aggarwal dan putri, penggunaan Adaptive Boost dapat digunakan bersama dengan mengkombinasikan algoritma klasifikasi lainnya untuk meningkatkan kinerja klasifikasi. Dalam upaya meningkatkan kinerja, setiap iterasi latihan memberikan bobot pada data dan bobot tersebut digunakan untuk melatih klasifikasi yang berbeda. Bobot akan diubah secara berulang kali berdasarkan kinerja klasifikasi yang terjadi [4].

III. HASIL DAN PEMBAHASAN

Pada tahap ini bertujuan untuk mengevaluasi performa model AdaBoostClassifier yang telah dilatih sebelumnya pada data uji. Langkah pertama adalah membuat prediksi menggunakan model tersebut dengan memanfaatkan fungsi predict, di mana data uji (x_test) digunakan sebagai input, dan hasil prediksi (y_pred) berupa label prediksi untuk setiap sampel dalam data uji. Prediksi ini merepresentasikan bagaimana model memetakan fitur dalam data uji ke label target yang telah didefinisikan. Prediksi ini merepresentasikan bagaimana model memetakan fitur dalam data uji ke label target yang telah didefinisikan.

Setelah prediksi selesai, evaluasi dilakukan dengan menghitung akurasi menggunakan fungsi accuracy_score. Akurasi adalah metrik dasar yang mengukur persentase prediksi yang benar terhadap total jumlah data uji. Nilainya ditampilkan dalam format desimal dua angka di belakang koma, misalnya Accuracy: 0.95, yang berarti model memiliki tingkat keakuratan 95%. Selanjutnya, evaluasi diperluas dengan menampilkan classification report menggunakan fungsi classification_report.

A. Dececion Tree

	precision	recall	f1-score	support
DBD	0.67	0.95	0.78	21
Gastroenteritis	0.93	0.82	0.87	49
ISK	1.00	0.88	0.93	8
ISPA	1.00	0.93	0.96	103
Thypoid	0.64	0.80	0.71	20
accuracy			0.89	201
macro avg	0.85	0.88	0.85	201
weighted avg	0.91	0.89	0.90	201

Gambar 4. Hasil Evaluasi Decicion Tree

Berdasarkan hasil evaluasi performa model berikut hal yang dapat dirangkum:

DBD:

-Precision: 0.67, menunjukkan bahwa dari semua prediksi yang mengindikasikan "DBD", 67% di antaranya benar.

-Recall: 0.95, menunjukkan bahwa model berhasil mendeteksi 95% kasus sebenarnya dari "DBD".

-F1-Score: 0.78, memberikan keseimbangan antara precision dan recall.

-Support: 21, menunjukkan bahwa terdapat 21 sampel sebenarnya dari kelas "DBD".

Gastroenteritis:

-Precision: 0.93, menunjukkan bahwa 93% prediksi untuk "Gastroenteritis" adalah benar.
-Recall: 0.82, artinya 82% dari kasus sebenarnya "Gastroenteritis" berhasil terdeteksi.
-F1-Score: 0.87, menyeimbangkan kedua metrik di atas.
-Support: 49, menandakan ada 49 sampel sebenarnya dari kelas ini.
ISK:
-Precision: 1.00, menunjukkan bahwa semua prediksi untuk kelas "ISK" benar.
-Recall: 0.88, menunjukkan bahwa model berhasil mendeteksi 88% dari kasus sebenarnya.
-F1-Score: 0.93, sangat baik dan stabil.
-Support: 8, menunjukkan jumlah sampel sebenarnya yang kecil.
ISPA:
-Precision: 1.00, menunjukkan bahwa semua prediksi untuk "ISPA" benar.
-Recall: 0.93, menunjukkan bahwa model mendeteksi 93% dari kasus sebenarnya.
-F1-Score: 0.96, menunjukkan performa sangat baik.
-Support: 103, jumlah terbesar dalam dataset.
Thypoid:
-Precision: 0.64, menunjukkan hanya 64% prediksi untuk "Thypoid" yang benar.
-Recall: 0.80, menunjukkan model mendeteksi 80% dari kasus sebenarnya.
-F1-Score: 0.71, memberikan performa yang cukup baik.
-Support: 20, menunjukkan ada 20 sampel sebenarnya.

Secara keseluruhan, model memiliki akurasi yang tinggi (89%) dan performa yang baik pada kelas mayoritas seperti "ISPA" dan "Gastroenteritis". Namun, untuk kelas minoritas seperti "DBD" dan "Thypoid", terutama "Thypoid", precision relatif lebih rendah, menunjukkan kebutuhan untuk peningkatan kemampuan model dalam menangani sampel dengan distribusi yang tidak merata. Algoritma SMOTE yang digunakan selama pelatihan membantu memperbaiki ketidakseimbangan kelas, tetapi beberapa kelas tetap memerlukan perhatian lebih dalam meningkatkan precision dan recall.

B. Support Vector Clasifer

	precision	recall	f1-score	support
DBD	0.53	0.95	0.68	21
Gastroenteritis	0.93	0.86	0.89	49
ISK	1.00	0.88	0.93	8
ISPA	1.00	0.92	0.96	103
Thypoid	0.94	0.75	0.83	20
accuracy			0.89	201
macro avg	0.88	0.87	0.86	201
weighted avg	0.93	0.89	0.90	201

Gambar 5. Hasil Evaluasi SVC

Berdasarkan hasil evaluasi performa model berikut hal yang dapat dirangkum.
DBD:
-Precision: 0.53, menunjukkan hanya 53% dari prediksi "DBD" yang benar.
-Recall: 0.95, menunjukkan bahwa model berhasil mendeteksi 95% kasus sebenarnya dari "DBD".
-F1-Score: 0.68, mencerminkan ketidakseimbangan antara precision yang rendah dan recall yang tinggi.
-Support: 21, menunjukkan ada 21 sampel sebenarnya dari kelas "DBD".
Gastroenteritis:
-Precision: 0.93, menunjukkan bahwa 93% prediksi untuk "Gastroenteritis" benar.
-Recall: 0.86, menunjukkan bahwa model mendeteksi 86% dari sampel sebenarnya.
-F1-Score: 0.89, menyeimbangkan antara precision dan recall.
-Support: 49, jumlah sampel sebenarnya dari kelas ini.
ISK:
-Precision: 1.00, menunjukkan bahwa semua prediksi untuk "ISK" benar.
-Recall: 0.88, menunjukkan bahwa model mendeteksi 88% dari kasus sebenarnya.
-F1-Score: 0.93, mencerminkan performa yang sangat baik.
-Support: 8, jumlah sampel sebenarnya yang kecil.
ISPA:
-Precision: 1.00, menunjukkan bahwa semua prediksi untuk "ISPA" benar.
-Recall: 0.92, menunjukkan bahwa model mendeteksi 92% dari sampel sebenarnya.
-F1-Score: 0.96, mencerminkan performa yang sangat kuat.
-Support: 103, jumlah sampel terbesar dalam dataset.
-Thypoid:
-Precision: 0.94, menunjukkan bahwa 94% dari prediksi "Thypoid" benar.
-Recall: 0.75, menunjukkan bahwa model mendeteksi 75% dari kasus sebenarnya.
-F1-Score: 0.83, mencerminkan performa yang cukup baik.

-Support: 20, jumlah sampel sebenarnya dari kelas ini.

- Model memiliki performa yang sangat baik secara keseluruhan, dengan akurasi sebesar 89% dan metrik rata-rata yang tinggi untuk semua kelas.

Model bekerja sangat baik pada kelas mayoritas seperti ISPA dan Gastroenteritis, dengan precision dan recall yang tinggi. Namun, untuk kelas minoritas seperti DBD, precision relatif rendah meskipun recall tinggi, menunjukkan bahwa model cenderung salah memprediksi kelas ini sebagai kelas lain. Sementara itu, performa untuk Thypoid juga cukup baik, tetapi recall yang lebih rendah mengindikasikan beberapa sampel yang sulit dideteksi. Dengan perbaikan pada balancing data atau tuning hyperparameter, performa untuk kelas minoritas dapat ditingkatkan lebih lanjut.

C. Logistic Regretion

	precision	recall	f1-score	support
DBD	0.72	0.86	0.78	21
Gastroenteritis	0.92	0.96	0.94	49
ISK	1.00	1.00	1.00	8
ISPA	1.00	0.93	0.96	103
Thypoid	0.81	0.85	0.83	20
accuracy			0.93	201
macro avg	0.89	0.92	0.90	201
weighted avg	0.93	0.93	0.93	201

Gambar 6. Hasil Evaluasi Logistic Regretion

Berdasarkan hasil evaluasi performa model berikut hal yang dapat dirangkum:
DBD:
-Precision: 0.72, yang berarti bahwa 72% dari prediksi untuk "DBD" adalah benar.
-Recall: 0.86, yang menunjukkan model berhasil mendeteksi 86% dari kasus sebenarnya untuk "DBD".
-F1-Score: 0.78, yang merupakan keseimbangan antara precision dan recall.
-Support: 21, menunjukkan ada 21 sampel sebenarnya dari kelas "DBD".
Gastroenteritis:
-Precision: 0.92, menunjukkan bahwa 92% dari prediksi untuk "Gastroenteritis" benar.
-Recall: 0.96, artinya 96% dari kasus sebenarnya "Gastroenteritis" terdeteksi dengan benar.
-F1-Score: 0.94, menunjukkan keseimbangan yang sangat baik antara precision dan recall.
-Support: 49, jumlah sampel yang relatif besar.
ISK:
-Precision: 1.00, menunjukkan bahwa semua prediksi untuk "ISK" adalah benar.
-Recall: 1.00, menunjukkan bahwa model berhasil mendeteksi semua sampel sebenarnya dari "ISK".
-F1-Score: 1.00, yang menunjukkan performa sempurna dalam prediksi kelas ini.
-Support: 8, meskipun jumlah sampel kecil, model sangat akurat dalam mendeteksi kelas ini.
ISPA:
-Precision: 1.00, menunjukkan semua prediksi untuk "ISPA" benar.
-Recall: 0.93, menunjukkan bahwa 93% dari kasus sebenarnya "ISPA" terdeteksi.
-F1-Score: 0.96, yang menunjukkan kinerja sangat baik dalam kelas ini.
-Support: 103, jumlah sampel terbesar dalam dataset.
Thypoid:
-Precision: 0.81, menunjukkan bahwa 81% dari prediksi untuk "Thypoid" adalah benar.
-Recall: 0.85, menunjukkan bahwa model berhasil mendeteksi 85% dari kasus sebenarnya.
-F1-Score: 0.83, memberikan performa yang cukup baik.
-Support: 20, menunjukkan jumlah sampel sebenarnya dari kelas ini.

Model menunjukkan performa yang sangat baik secara keseluruhan dengan akurasi 93% dan metrik evaluasi yang tinggi. Kelas-kelas utama seperti ISK, ISPA, dan Gastroenteritis mendapat hasil yang sangat baik dengan precision dan recall yang tinggi, bahkan mencapai performa sempurna di beberapa kelas. Meskipun ada sedikit ruang untuk perbaikan pada kelas DBD dan Thypoid, model masih menunjukkan hasil yang solid pada kelas-kelas tersebut, dengan recall yang cukup tinggi. Secara keseluruhan, model ini dapat diandalkan untuk prediksi penyakit dengan akurasi dan keseimbangan yang sangat baik antar kelas.

IV. KESIMPULAN

1. Berdasarkan hasil pengujian menggunakan tiga metode Base Estimator yang berbeda dalam model Adaptive Boost, dapat disimpulkan bahwa jenis Base Estimator memiliki pengaruh signifikan terhadap performa model dalam prediksi penyakit akut. Hasil pengujian menunjukkan bahwa Logistic Regression memiliki performa terbaik dibandingkan dengan Support Vector Classifier (SVC) dan Decision Tree, dengan akurasi keseluruhan sebesar 93%, serta nilai precision dan recall yang lebih stabil untuk berbagai jenis penyakit.
2. Pemilihan Base Estimator yang lebih kompleks tidak selalu menjamin peningkatan akurasi. Meskipun metode seperti SVC dan Decision Tree memiliki keunggulan dalam beberapa aspek, hasil pengujian menunjukkan bahwa Logistic Regression yang relatif lebih sederhana justru memberikan hasil

yang lebih baik secara keseluruhan. Hal ini terlihat dari peningkatan akurasi dan keseimbangan antara precision dan recall pada hampir semua kategori penyakit. Dengan demikian, pemilihan Base Estimator yang optimal harus mempertimbangkan keseimbangan antara kompleksitas model dan kinerjanya pada dataset tertentu, bukan hanya berdasarkan tingkat kompleksitasnya.

REFERENSI

[1] L. D. Andhika, D. R. Cahyani, D. Saputra, and T. Herawati, “Analisis Sentimen Kosumen KFC Berdasarkan Pendekatan Naive Bayes dan Ada Boost Berbasis Data Twitter,” vol. 3, no. 1, pp. 55–61, 2023.

[2] A. Desiani, I. Maiyanti, B. Suprihatin, R. Kurniawan, D. Adzra, A. Nabila, J. Matematika, F. Matematika, dan I. P. Alam, "Implementasi algoritma Adaptive Boosting (Adaboost) dan Single Layer Perceptron (SLP) pada klasifikasi penyakit Hepatitis-C," ScientiCO: Computer Science and Informatics Journal, vol. 6, no. 2, pp. 45–52, 2023.

[3] A. Irma Prianti, “Pebandingan Metode K-Nearest Neighbor dan Adaptive Boosting pada Kasus Klasifikasi Multi Kelas,” J Stat. J. Ilm. Teor. dan Apl. Stat., vol. 13, no. 1, pp. 39–47, 2020, doi: 10.36456/jstat.vol13.no1.a3269.

[4] T. A. E. Putri, T. Widiharih, dan R. Santoso, "Penerapan tuning hyperparameter RandomSearchCV pada Adaptive Boosting untuk prediksi kelangsungan hidup pasien gagal jantung," Jurnal Gaussian, vol. 11, no. 3, pp. 397–406, 2023.