

Perbandingan Algoritma Real Adaptive Boosting pada Regresi Logistik, CART, dan Naive Bayes dalam Klasifikasi Biji Labu

Moch. Anjas Aprihartha¹, Sefri Imanuel Fallo², Hady Rasikhun³

¹Program Studi PJJ Informatika, Universitas Dian Nuswantoro
Jl. Imam Bonjol No.207, Pendrikan Kidul, Kota Semarang, Jawa Tengah 50131, Indonesia

²Program Studi Matematika, Universitas San Pendro

Jl. Ir. Soekarno, Fontein, Kota Kupang, Nusa Tenggara Timur, Indonesia

³Program Studi Teknik Pertambangan, Universitas Muhammadiyah Mataram
Jl. KH. Ahmad Dahlan No.1, Pagesangan, Kota Mataram, Nusa Tenggara Barat, Indonesia

Email: anjas.aprihartha@dsn.dinus.ac.id¹, fallosefrimanuel@gmail.com²,

hady.rasikhun@ummat.ac.id³

Korespondensi penulis: anjas.aprihartha@dsn.dinus.ac.id¹

Submitted : 16 Mei 2025

Accepted : 30 Juli 2025

Published : 30 Juli 2025

Abstrak

Labu merupakan spesies tanaman yang bernilai ekonomis dan medis. Hampir setiap bagian dari labu dapat dikonsumsi terutama pada bijinya. Minyak dari biji labu dapat juga digunakan sebagai saus untuk salad, produk kosmetik, sabun dan lilin. Keterampilan dalam mengklasifikasikan biji labu dengan tepat sangat dibutuhkan diberbagai sektor, seperti pertanian dan industri pangan. Dibutuhkan teknologi pengembangan yang dapat mengidentifikasi dan mensortir biji labu dengan mudah dan cepat. Beberapa algoritma yang umum dapat digunakan untuk mengidentifikasi jenis biji labu seperti algoritma regresi logistik (RL), *Classification and Regression Tree* (CART), dan Naive Bayes (NB). Penelitian ini bertujuan mengeksplorasi model RL, CART, dan NB pada dua jenis varietas biji labu, yaitu Ürgüp Sivrisi dan Çerçevelik berdasarkan karakteristik fisiknya. Selain itu, digunakan pendekatan *Real Adaptive Boosting* (RAB) untuk meningkatkan kinerja model dasar. Teknik ini bekerja dengan kemampuan menggabungkan beberapa model homogen secara berulang untuk menghasilkan model yang kuat. Hasil uji kinerja model klasifikasi diperhitungkan melalui metrik evaluasi. Model RAB-RL memiliki performa tertinggi pada akurasi, presisi, dan *f1-score* sehingga menjadikan model terbaik dalam mengklasifikasikan jenis biji labu dibandingkan model-model lainnya. Dalam model dasar, model RL memiliki performa terbaik dibawah model RAB-RL.

Kata Kunci: biji labu, CART, Naive Bayes, real adaptive boosting, regresi logistik.

Abstract

Pumpkin is a plant species that has economic and medical value. Almost every part of the pumpkin can be consumed, especially the seeds. Pumpkin seed oil can also be used as a salad

dressings, cosmetic products, soaps and candles. The skill of classifying pumpkin seeds correctly is needed in various sectors, such as agriculture and the food industry. It is necessary to develop technology that can identify and sort pumpkin seeds easily and quickly. Several common algorithms can be used to identify pumpkin seed types such as logistic regression (RL), Classification and Regression Tree (CART), and Naive Bayes (NB) algorithms. This study aims to explore the RL, CART, and NB models on two types of pumpkin seed varieties, namely Ürgüp Sivrisi and Çerçvelik based on their physical characteristics. In addition, the Real Adaptive Boosting (RAB) approach is used to improve the performance of the base model. This technique works with the ability of several homogeneous models repeatedly to produce a robust model. The results of the classification model performance test are calculated through evaluation metrics. The RAB-RL model has the highest performance in accuracy, precision, and f1-score, making it the best model in classifying pumpkin seed types compared to other models. In the baseline model, the RL model has the best performance under the RAB-RL model.

Keywords: CART, logistic regression, Naive Bayes, pumpkin seed, real adaptive boosting.

1. Pendahuluan

Labu merupakan spesies tanaman yang bernilai ekonomis dan medis [1]. Bagian daging dari labu digunakan dalam sup, hidangan sayur, kue, makanan penutup, dan penganan manis. Sementara bijinya dikonsumsi sebagai makanan ringan, bagian limbahnya digunakan dalam nutrisi hewan. Selain dikonsumsi segar atau dipanggang, Minyak dari biji labu dapat juga digunakan sebagai suplemen saus untuk salad, produk kosmetik, sabun dan lilin. Biji labu dan minyak biji labu dipercaya memiliki banyak manfaat kesehatan karena komposisi makro dan mikronya. Biji labu kaya akan antioksidan, zat besi, seng, magnesium, dan banyak nutrisi lain seperti karotenoid dan vitamin E [2]. Selain 37% karbohidrat, biji labu juga mengandung 35% hingga 40% lemak, dan protein. Labu memiliki berbagai jenis, Ürgüp Sivrisi merupakan salah satu spesiesnya. Jenis biji labu yang dikenal sebagai Ürgüp Sivrisi memiliki cangkang yang panjang, putih, sangat cerah, tipis, dan hampir tidak terlihat dengan ujung yang runcing [3]. "Çerçvelik" adalah varietas biji labu lainnya. Spesies khusus ini disebut "Topak" yang tumbuh di Nevşehir, Karacaören, Turki.

Pada masa ini, keterampilan dalam mengklasifikasikan biji labu dengan tepat sangat dibutuhkan diberbagai sektor, seperti pertanian dan industri pangan [4]. Karakteristik bentuk dan ukuran benih dipertimbangkan dalam desain mesin sortasi dan penggilingan. Atribut fisik benih labu harus diketahui untuk desain peralatan yang akan membantu dari penanaman benih hingga pemrosesan pasca panen dan pemasaran. Untuk mengatasi masalah ini, diperlukan teknologi pengembangan yang dapat mengidentifikasi dan mensortir biji labu dengan mudah dan cepat.

Beberapa algoritma klasifikasi yang umum dapat digunakan seperti regresi logistik, *Classification and Regression Tree* (CART), dan Naive Bayes dapat digunakan dalam mengidentifikasi biji labu dengan cepat. Penelitian oleh Koklu, *et al.* [5] yang menggunakan algoritma regresi logistik dalam mengklasifikasikan dua jenis biji labu yaitu Ürgüp Sivrisi dan Çerçvelik dengan mempertimbangkan fitur morfologinya. Kinerja model ditentukan dengan bantuan metode *10-fold cross validation*. Tingkat akurasi pengklasifikasi diperoleh sebesar 87,92 persen. Penelitian oleh Prasad, *et al.* [6] yang mengimplementasikan metode regresi logistik, Naive Bayes dan *decision tree* dalam memprediksi jenis biji labu. Hasil penelitian diperoleh metode regresi logistik, Naive

Bayes, dan *decision tree* masing-masing memberikan keakuratan prediksi sebesar 99,81%; 95,80%; dan 100%.

Dalam beberapa kasus, algoritma klasifikasi dasar belum cukup memberikan kinerja klasifikasi, tiap-tiap algoritma memiliki keterbatasan dalam mengolah data, seperti data yang tidak seimbang dan pemilihan atribut yang tidak sesuai. Dalam mengatasi masalah ini maka dibutuhkan pendekatan dengan algoritma bantu, yaitu algoritma *Real Adaptive Boosting* (RAB). Teknik ini bekerja dengan kemampuan menggabungkan beberapa model homogen secara berulang untuk menghasilkan model yang kuat. Penelitian yang dilakukan oleh Aprihartha, *et al.* [7] dalam membandingkan algoritma dasar CART dengan algoritma CART-RAB dalam memprediksi minat pelanggan membeli sepatu. Hasil penelitian diperoleh akurasi model CART sebesar 77,5% sedangkan model CART-RAB sebesar 85,71%. Sehingga model yang diusulkan dapat digunakan secara efektif adalah model CART-RAB.

Berdasarkan uraian yang telah diberikan, tujuan dari studi ini adalah mengeksplorasi model klasifikasi pada dua jenis varietas biji labu menurut karakteristik fisiknya. Penelitian ini akan berpusat pada perbandingan kinerja model dasar seperti regresi logistik, Naïve Bayes, dan CART dengan model RAB pada algoritma dasar dari regresi logistik, Naïve Bayes, dan CART. Model yang terbaik berdasarkan kinerjanya dapat digunakan secara efektif untuk karakterisasi, klasifikasi, dan identifikasi biji labu. Hasil penelitian ini diharapkan dapat menjadi acuan dalam pengembangan sistem seleksi biji labu untuk memperkuat efisiensi di industri.

2. Metode Penelitian

2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari kumpulan data repositori, Kaggle. Data awal merupakan dataset citra yang telah diekstraksi menjadi data numerik. Dataset ini terdiri dari 2500 amatan dengan dua kategori kelas pada variabel dependen, yaitu biji labu jenis Çerçevelik dan jenis Ürgüp Sivrisi. Sedangkan variabel independen terdiri dari 12 karakteristik morfologi biji labu yang meliputi: *Area, Perimeter, Major Axis Length, Minor Axis Length, Convex Area, Equiv Diameter, Eccentricity, Solidity, Extent, Roundness, Aspect Ration, Compactness*.

2.2 Metode Regresi Logistik

Metode statistik yang umum digunakan dalam klasifikasi biner adalah regresi logistik [8]. Metode regresi logistik merupakan metode analisis yang dikembangkan dari metode regresi linear dengan variabel depeden berbentuk kategorikal. Metode ini diterapkan untuk memprediksi probabilitas kelas biner berdasarkan variabel independen dengan mengikuti distribusi Bernaulli. Regresi logistik menggunakan *odds ratio* untuk mewakili probabilitas peristiwa yang terjadi (p) dengan probabilitas peristiwa yang tidak terjadi ($1 - p$) [9].

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (1)$$

Probabilitas posterior dari konvensional regresi logistik dijelaskan sebagai berikut:

$$p = \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{1}{1+e^{-g(x)}} \quad (2)$$

dengan $g(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$, Koefisien β_0 dan β_i merupakan parameter estimasi untuk intersep dan variabel independen X_i dengan $i = 1, 2, \dots, m$.

2.3 Metode Naive Bayes

Metode Naive Bayes merupakan algoritma klasifikasi yang paling representatif diantara algoritma klasifikasi jaringan bayessian lainnya, dengan asumsi semua atribut dalam kelas tertentu saling independen [10]. Misalkan diberikan himpunan data $D = \{X_1, X_2, \dots, X_m\}$. Setiap amatan X_i direpresentasikan oleh vektor atribut berdimensi N , $\langle a_{i1}, a_{i2}, a_{i3}, \dots, a_{iN} \rangle$. Metode Naive Bayes memprediksi label kelas X_i menggunakan persamaan berikut:

$$c(X_i) = \operatorname{argmax}_{c \in C} \frac{P(c) \prod_{j=1}^N P(a_{ij}|c)}{\sum_{c \in C} P(c) \prod_{j=1}^N P(a_{ij}|c)} \quad (3)$$

dengan N melambangkan jumlah atribut, C adalah himpunan semua kemungkinan label kelas c , dan a_{ij} merujuk kepada nilai variabel independen ke- j dari amatan ke- i .

Probabilitas prior pada kelas c dinyatakan dalam persamaan berikut:

$$P(c) = \frac{\sum_{r=1}^m I(c_r, c) + 1}{m + \vartheta(C)} \quad (4)$$

Probabilitas a_{ij} bersyarat c dapat dihitung dengan persamaan berikut:

$$P(a_{ij}|c) = \frac{\sum_{r=1}^m I(c_r, c) I(a_{ij}, a_{rj}) + 1}{\sum_{r=1}^m I(c_r, c) + \vartheta(A_j)} \quad (5)$$

dengan A_j mengacu pada variabel independen ke- j , dan c_r mewakili label kelas dari amatan ke- r di dataset training. Fungsi $\vartheta(\sim)$ digunakan untuk menghitung jumlah nilai unik di C atau A_j . Selain itu, $I(\sim)$ merupakan indikator fungsi, memberikan nilai 1 ketika kondisi internalnya terpenuhi dan 0 sebaliknya.

2.4 Metode Classification and Regression Tree

Classification and Regression Tree (CART) merupakan pendekatan *supervised learning* yang berguna untuk mengelompokkan ruang prediktor/fitur ke dalam wilayah yang lebih kecil dengan hasil yang lebih homogen [11]. Kumpulan segmen ini diringkas dalam pohon keputusan. Salah satu keuntungan dari pendekatan berstruktur pohon adalah melakukan pemilihan variabel secara otomatis dan dapat menganalisis berbagai jenis fitur, terlepas dari sifat atau distribusinya. Selain itu, CART mudah diinterpretasikan melalui pembacaan visual yang jelas.

Model CART dibangun secara berulang dengan mempartisi data berdasarkan nilai *Gini Index* terbesar [12]. Misalkan diberikan himpunan data S , *Gini Index* dinyatakan dalam persamaan berikut [13].

$$Gini(S) = 1 - \sum_{i=0}^{c-1} p_i^2 \quad (6)$$

dengan $p_i = \frac{s_i}{S}$ adalah frekuensi relatif kelompok C_i dalam himpunan, s_i adalah banyaknya sampel pada kelompok C_i , dan C_i merupakan kelompok/ kelas untuk $i = 0, 1, 2, \dots, c - 1$.

2.5 Metode Real Adaptive Boosting

Real Adaptive Boosting (RAB) merupakan algoritma pembelajaran gabungan yang bertujuan untuk membangun model klasifikasi kuat dari beberapa model klasifikasi lemah dalam meningkatkan kinerja model [7]. Algoritma ini menghasilkan nilai probabilitas yang mencerminkan keyakinan model dalam menentukan kelas pada data baru.

Langkah-langkah algoritma RAB sebagai berikut:

1. Diawali dengan menetapkan bobot untuk dataset training, $w_i = \frac{1}{N}, i = 1, 2, \dots, N$.
2. Lakukan iterasi untuk $m = 1, 2, \dots, M$.
 - a. Bangun model klasifikasi $f_m(x)$ dengan menggunakan bobot w_i .

- b. Menghitung $F_M(x_i) = \sum_{l=1}^m f_l(x_i)$.
- c. Menghitung probabilitas setiap amatan ke- i , $p_m(x_i) = \frac{1}{1+e^{-2F_M(x_i)}}$ dengan $p_m(x) = \hat{P}_w(y = 1|x) \in [0,1]$.
- d. Perbaharui bobot $w_i \leftarrow w_i e^{(-y_i F_M(x_i))}$, $i = 1, 2, \dots, N$.
3. Menguji model klasifikasi dengan data testing.
4. Menghitung kinerja model klasifikasi.

2.6 Uji Kinerja Model Klasifikasi

Metrik yang digunakan dalam menguji kemampuan model klasifikasi diantaranya akurasi, presisi, *recall*, dan *f1-score* yang didefinisikan dalam persamaan berikut [14] [15].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (9)$$

$$F1 - score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \times 100\% \quad (10)$$

dengan di mana TP merupakan positif benar, TN merupakan negatif benar, FN merupakan negatif benar, dan FP merupakan positif salah [16].

Akurasi dinyatakan sebagai tingkat keberhasilan model dalam mengidentifikasi titik data dengan tepat [17] [18]. Presisi dan *recall* mengukur kemampuan model klasifikasi untuk mengidentifikasi hanya titik data yang relevan dan menemukan semua kasus yang relevan dalam kumpulan data. Presisi yang tinggi menunjukkan rasio positif palsu yang rendah, sedangkan *recall* yang tinggi menunjukkan rasio negatif palsu yang rendah. *F1-score* adalah ukuran gabungan dari presisi dan *recall*. *F1-score* yang tinggi menyiratkan model klasifikasi yang kuat [19].

3. Hasil dan Pembahasan

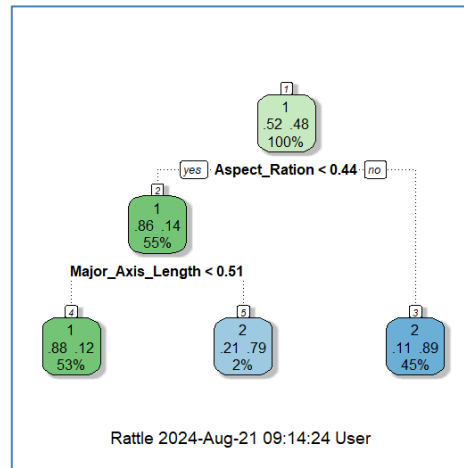
Sebelum membuat model klasifikasi, dataset akan diproses dengan tahap transformasi untuk menyelaraskan jangkauan data setiap variabel independen. Transformasi yang digunakan adalah transformasi *Z-Score Normalization* yang dinyatakan dalam persamaan berikut:

$$Z = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

dengan x_{ij} merupakan observasi ke- i pada variabel ke- j , $\bar{x}_j$ merupakan rata-rata pada variabel ke- j , dan σ_j merupakan deviasi standar pada variabel ke- j .

Setelah data ditransformasi, selanjutnya dataset dibagi menjadi dua yaitu dataset *training* dan dataset *testing* dengan perbandingan 4:1. Proses pembentukan dataset *training* dilakukan dengan teknik *simple random sampling* sehingga setiap amatan memiliki kesempatan yang sama untuk masuk kedalam dataset *training* sedangkan sisanya yang tidak terpilih masuk kedalam dataset *testing*. Dataset awal terdiri 2500 amatan yang dialokasikan menjadi dataset *training* sebanyak 2000 amatan dan sisanya 500 amatan membentuk dataset *testing*.

Dataset *training* digunakan untuk melatih algoritma sehingga menghasilkan model. Algoritma Naive Bayes (NB) menghasilkan model berupa distribusi probabilitas bersyarat untuk setiap kelas berdasarkan variabel-variabel independen. Algoritma CART menghasilkan model dalam bentuk diagram pohon keputusan seperti pada Gambar 1.



Gambar 1. Model CART

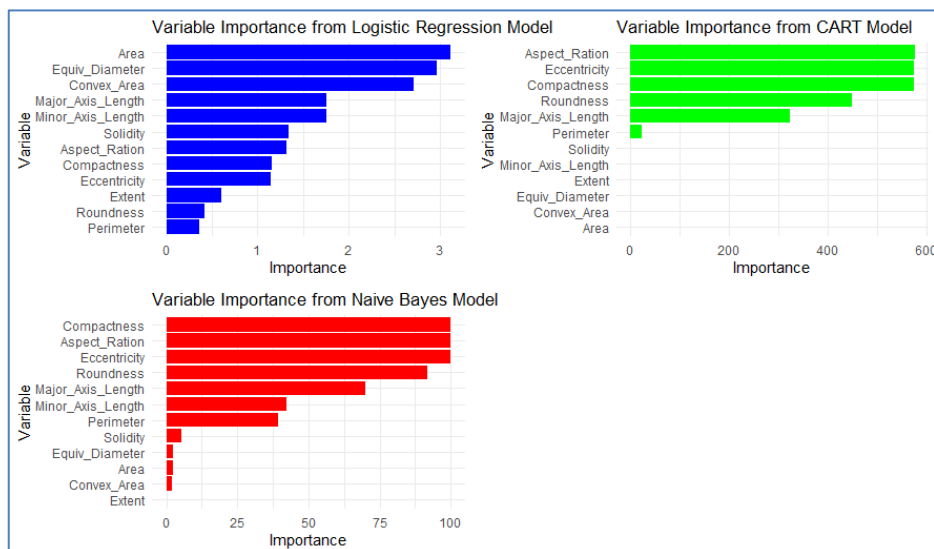
Pada Gambar 1, variabel yang menjadi simpul akar adalah *Aspect Ration*. Apabila terdapat sampel baru dengan *Aspect Ration* $\geq 0,44$ maka diklasifikasikan kedalam biji labu jenis Ürgüp Sivrisi. Sebaliknya jika *Aspect Ration* $< 0,44$ maka dilanjutkan pemisahan berdasarkan *Major Axis Length*. Jika *Major Axis Length* $\geq 0,51$ maka diklasifikasikan kedalam biji labu Ürgüp Sivrisi, sebaliknya jika *Major Axis Length* $< 0,51$ maka diklasifikasikan kedalam biji labu Çerçvelik.

Berbeda dengan algoritma Naïve Bayes dan CART, algoritma Regresi Logistik (RL) menghasilkan model berupa persamaan dalam *log odds* (logit) seperti berikut:

$$\pi(x) = \frac{\exp \left(\frac{93,91+730,04Area-19,86Perimeter +102,23Major Axis Length+67,77Minor Axis Length-646,59Convex Area-191,27Equiv Diameter-49,20Eccentricity -21,58Solidity+0,33Extent-9,19Roundness -33,71Aspect Ration-68,52Compactness}{93,91+730,04Area-19,86Perimeter +102,23Major Axis Length+67,77Minor Axis Length-646,59Convex Area-191,27Equiv Diameter-49,20Eccentricity -21,58Solidity+0,33Extent-9,19Roundness -33,71Aspect Ration-68,52Compactness} \right)}{1+\exp \left(\frac{93,91+730,04Area-19,86Perimeter +102,23Major Axis Length+67,77Minor Axis Length-646,59Convex Area-191,27Equiv Diameter-49,20Eccentricity -21,58Solidity+0,33Extent-9,19Roundness -33,71Aspect Ration-68,52Compactness}{93,91+730,04Area-19,86Perimeter +102,23Major Axis Length+67,77Minor Axis Length-646,59Convex Area-191,27Equiv Diameter-49,20Eccentricity -21,58Solidity+0,33Extent-9,19Roundness -33,71Aspect Ration-68,52Compactness} \right)}$$

dengan nilai $0 \leq \pi(x) \leq 1$. Jika sampel baru berada pada $\pi(x) < 0,5$ maka sampel tersebut diklasifikasikan kedalam biji labu Çerçvelik, sebaliknya jika $\pi(x) \geq 0,5$ maka sampel berada dalam klasifikasi biji labu Ürgüp Sivrisi.

Proses selanjutnya mengestimasi seberapa penting setiap variabel independen berpengaruh terhadap klasifikasi biji labu dalam algoritma RL, CART, dan NB. Hasil uji ditampilkan pada Gambar 2.



Gambar 2. Variabel Signifikan Model

Pada model RL, variabel *Area* memberikan pengaruh signifikan terhadap pembentukan model klasifikasi, disusul oleh variabel *Equiv Diameter* dan *Convex Area*. Sementara itu pada model CART, variabel *Aspect Ration* memberikan kontribusi terbesar dalam proses klasifikasi biji labu. Sedangkan variabel *Solidity*, *Minor Axis Length*, *Exent*, *Equiv Diameter*, *Convex Area*, *Area* tidak berperan penting dalam proses prediksi dan dianggap tidak relevan dalam pembentukan model CART. Dalam algoritma NB, variabel *Compactness*, *Aspect Ration*, dan *Eccentricity* sama-sama memiliki kontribusi penting dalam klasifikasi biji labu. Sebaliknya variabel *Extent* berada diposisi paling bawah dan tidak memiliki pengaruh kuat dalam klasifikasi biji labu.

Algoritma dasar seperti regresi logistik (RL), CART, dan Naive Bayes (NB) diintegrasikan dengan *ensemble learning* yaitu *Real Adaptive Boosting* (RAB). Teknik ini menciptakan kumpulan model sederhana berdasarkan model dasar yang telah ditetapkan. Setiap model yang sudah dilatih kemudian diuji dengan dataset testing kemudian diukur melalui lima metrik evaluasi. Hasil uji performa setiap model klasifikasi disajikan dalam Tabel 1.

Tabel 1. Kinerja Model Klasifikasi

Ukuran	RL	NB	CART	RAB-RL	RAB-NB	RAB-CART
Akurasi	87.80%	85.20%	85.80%	88.20%	85.40%	87.00%
Presisi	92.25%	86.82%	86.82%	92.25%	87.21%	87.21%
Recall	85.30%	84.85%	85.82%	85.92%	84.91%	87.55%
F1-score	88.64%	85.82%	86.32%	88.97%	86.04%	87.38%
Rata-rata	88.50%	85.67%	86.19%	88.84%	85.89%	87.29%

Salah satu kelemahan model *Real Adaptive Boosting* adalah waktu komputasinya yang cenderung lebih lama dibandingkan dengan model klasifikasi tunggal, karena proses pelatihan yang melibatkan sejumlah model lemah secara berurutan. Namun, model ini umumnya memiliki kinerja yang lebih baik dibandingkan model tunggal, karena kemampuannya dalam menggabungkan beberapa model lemah untuk meningkatkan akurasi prediksi. Berdasarkan Tabel 1, Model RL memiliki rata-rata kinerja sebesar 88,50%, sedangkan model RAB-RL menunjukkan peningkatan kinerja rata-rata menjadi

88,84%. Sehingga, terdapat peningkatan sebesar 0,34%. Selain itu, integrasi model tunggal NB dan CART ke dalam kerangka *Real Adaptive Boosting* menghasilkan peningkatan performa masing-masing sebesar 0,22% dan 1,10%. Di antara ketiga model, peningkatan terbesar dicapai oleh model CART.

Model *Real Adaptive Boosting*-Regresi Logistik (RAB-RL) memiliki akurasi tertinggi (88,20%) sedangkan model NB memiliki akurasi terendah (85,20%). Persentase presisi tertinggi dimiliki oleh model RL dan RAB-RL. Model *Real Adaptive Boosting*-CART (RAB-CART) memiliki *recall* lebih baik (87,55%) dibandingkan model lainnya. Model RAB-RL memiliki *f1-score* sebesar 88,97%, lebih baik dibandingkan lima model lainnya. Selanjutnya dilakukan uji Friedman dalam mengidentifikasi apakah terdapat perbedaan signifikan antara keenam model tersebut. Hasil uji Friedman disajikan dalam Tabel 2.

Tabel 2. Hasil Uji Friedman

Friendman chi-square	16,168
df	5
<i>p-value</i>	0,006381

Berdasarkan Tabel 2, dengan menetapkan $\alpha = 0,05$ diperoleh bahwa nilai *p-value* ($0,006381 < \alpha(0,05)$). Hal ini mengidentifikasikan bahwa terdapat perbedaan yang signifikan antara keenam model klasifikasi. Dengan demikian, model *Real Adaptive Boosting*-Regresi Logistik (RAB-RL) memiliki performa tertinggi hampir di semua metrik evaluasi dengan rata-rata sebesar 88,84%. Model ini memberikan akurasi paling baik dalam klasifikasi jenis biji labu. Disusul oleh model Regresi Logistik (RL) yang memiliki performa terbaik dibawah model RAB-RL. Sementara itu, model Naive Bayes (NB) memiliki performa model terendah berdasarkan persentase metrik evaluasi.

4. Kesimpulan

Penelitian ini menerapkan teknik klasifikasi dengan enam jenis algoritma pada proses klasifikasi biji labu, diantaranya algoritma regresi logistik, Naive Bayes, *Classification and Regression Tree* (CART), *Real Adaptive Boosting*-Regresi Logistik, *Real Adaptive Boosting*-Naive Bayes, *Real Adaptive Boosting*-CART. Pada model Regresi Logistik, variabel *Area* memberikan pengaruh signifikan terhadap pembentukan model klasifikasi, disusul oleh variabel *Equiv Diameter* dan *Convex Area*. Pada model CART, variabel *Aspect Ration* memberikan kontribusi terbesar dalam proses klasifikasi biji labu. Sedangkan pada model Naive Bayes, variabel *Compactness*, *Aspect Ration*, dan *Eccentricity* bersama-sama memiliki kontribusi penting dalam klasifikasi biji labu.

Hasil uji kemampuan model dalam klasifikasi diperhitungkan melalui metrik evaluasi. Integrasi model tunggal dari RL, NB, dan CART ke dalam kerangka *Real Adaptive Boosting* menghasilkan peningkatan performa masing-masing sebesar 0,34%, 0,22%, dan 1,10%. Di antara ketiga model, peningkatan terbesar dicapai oleh model CART.

Selain itu, hasil uji Friedman menunjukkan terdapat perbedaan yang signifikan antara keenam model klasifikasi yang ditunjukkan dengan nilai *p-value* ($0,006381 < \alpha(0,05)$). Model *Real Adaptive Boosting*-Regresi Logistik (RAB-RL) memiliki performa tertinggi pada akurasi, presisi, dan *f1-score* sehingga menjadikan model terbaik dalam mengklasifikasikan jenis biji labu dibandingkan model-model lainnya. Model ini memiliki performa tertinggi hampir di semua metrik evaluasi dengan rata-rata sebesar 88,84%. Dalam model tunggal, model Regresi Logistik memiliki performa terbaik dibawah model

Real Adaptive Boosting-Regresi Logistik. Sementara itu, model Naive Bayes (NB) memiliki performa model terendah berdasarkan persentase metrik evaluasi yakni sebesar 85,67%.

Daftar Pustaka

- [1] D. Montesano, G. Rocchetti, P. Putnik, and L. Lucini, "Bioactive profile of pumpkin: an overview on terpenoids and their health-promoting properties," *Current Opinion in Food Science*, vol. 22, pp. 81–87, Aug. 2018, doi: 10.1016/j.cofs.2018.02.003.
- [2] G. O. Fruhwirth, T. Wenzl, R. El-Toukhy, F. S. Wagner, and A. Hermetter, "Fluorescence screening of antioxidant capacity in pumpkin seed oils and other natural oils," *European Journal of Lipid Science and Technology*, vol. 105, no. 6, pp. 266–274, Jun. 2003, doi: 10.1002/ejlt.200390055.
- [3] N. Aktaş, T. Uzlaşır, and Y. E. Tunçil, "Pre-roasting treatments significantly impact thermal and kinetic characteristics of pumpkin seed oil," *Thermochimica Acta*, vol. 669, pp. 109–115, Nov. 2018, doi: 10.1016/j.tca.2018.09.012.
- [4] N. Çetin *et al.*, "Binary classification of pumpkin (*Cucurbita pepo* L.) seeds based on quality features using machine learning algorithms," *European Food Research and Technology*, vol. 250, no. 2, pp. 409–423, Feb. 2024, doi: 10.1007/s00217-023-04392-w.
- [5] M. Koklu, S. Sarigil, and O. Ozbek, "The use of machine learning methods in classification of pumpkin seeds (*Cucurbita pepo* L.)," *Genetic Resources and Crop Evolution*, vol. 68, no. 7, pp. 2713–2726, Oct. 2021, doi: 10.1007/s10722-021-01226-0.
- [6] K V Prasad, Hanumesh Vaidya, Kumar Swamy K, and Renuka S, "Pumpkin Seeds Classification: Artificial Neural Network and Machine Learning Methods", *J. Int. Acad. Phys. Sci.*, vol. 27, no. 1, pp. 23–33, Mar. 2023.
- [7] A. Aprihartha, J. Prasetya, and S. I. Fallo, "Implementasi CART-Real Adaboost dalam Memprediksi Minat Pelanggan Membeli Sepatu," *Jurnal EurekaMatika*, vol. 12, no. 1, pp. 35–46, 2024, doi.org/10.17509/jem.v12i1.67808.
- [8] J. Prasetya, S. I. Fallo, and M. A. Aprihartha, "Stacking Machine Learning Model for Predict Hotel Booking Cancellations," *Jurnal Matematika, Statistika dan Komputasi*, vol. 20, no. 3, pp. 525–537, May 2024, doi: 10.20956/j.v20i3.32619.
- [9] M. A. Aprihartha and I. Idham, "Optimization of Classification Algorithms Performance with k-Fold Cross Validation", *EMJ*, vol. 7, no. 2, pp. 61–66, Sep. 2024, doi.org/10.29303/emj.v7i2.212.
- [10] X. Zhou, Y. Wang, L. Zhang, A. Huang, and X. Wang, "An innovative multi-view collaborative optimization framework for Weighted Naive Bayes," *Knowledge-Based Systems*, vol. 317, p. 113378, May 2025, doi: 10.1016/j.knosys.2025.113378.
- [11] V. Calabrese *et al.*, "Review and practical excursus on the comparison between traditional statics methods and Classification And Regression Tree (CART) in real-life data: Low protein diet compared to Mediterranean diet in patients with chronic kidney disease," *Nefrología*, vol. 45, no. 4, pp. 279–284, Apr. 2025, doi: 10.1016/j.nefro.2025.01.009.
- [12] A. Aprihartha, Zulhandi Putrawan, Dicky Zulhan, and Fatma Ahardika Nurfaizal, "Klasifikasi Produktivitas Buah Nanas Menggunakan Algoritma

- Classification and Regression Tree (CART)," *Diophantine Journal of Mathematics and Its Applications*, pp. 64–70, Jul. 2024, doi: 10.33369/diophantine.v3i1.34193.
- [13] M. anjas Aprihartha, F. Astutik, and N. Sulistianingsih, "Comparison of Naïve Bayes, CART, dan CART Adaboost Methods in Predicting Tire Product Sales," *Jurnal Matematika, Statistika dan Komputasi*, vol. 20, no. 3, pp. 596–605, May 2024, doi: 10.20956/j.v20i3.33187.
- [14] M. anjas Aprihartha, Z. Putrawan, D. Zulhan, and F. A. Nurfaizal, "Study On Identification Of Poisonous and Non-Toxic Mushrooms Using the Cart-Logitboost Algorithm," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 33–45, Sep. 2024, doi: 10.20956/j.v21i1.35072.
- [15] Moch. A. Aprihartha, T. N. Alam, and M. Husniyadi, "Perbandingan Metrik Euclidean dan Metrik Manhattan untuk K-Nearest Neighbors dalam Klasifikasi Kismis," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 21–30, Jun. 2024, doi: 10.54082/jiki.126.
- [16] M. Anjas Aprihartha, Z. Putrawan, D. Zulhan, F. Ahardika Nurfaizal, and G. Artikel, "Algoritma Synthetic Minority Oversampling Technique dan C5.0 dalam Mengatasi Ketidakseimbangan Data pada Klasifikasi Kelulusan Siswa," *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, vol. 2, no. 1, pp. 1–10, 2024, doi: 10.30812/upgrade.v2i1.4148.
- [17] A. Aprihartha, "Penyelesaian Masalah Ketidakseimbangan Data Melalui Teknik Oversampling dan Undersampling pada Klasifikasi Siswa Tidak Naik Kelas," *Jurnal Teknik Ibnu Sina (JT-IBSI)*, vol. 9, no. 01, pp. 43–52, 2024, doi.org/10.36352/jt-ibsi.v9i01.807.
- [18] A. Aprihartha, "Implementasi Metode Support Vector Machine (SVM) Pada Klasifikasi Status Penerima Bantuan Pangan Non Tunai," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 16, no. 2, pp. 313–324, 2024, doi.org/10.18495/jsi.v16i2.123.
- [19] P. Tongcham, P. Supa, P. Pornwongthong, and P. Prasitmeeboon, "Mushroom spawn quality classification with machine learning," *Computers and Electronics in Agriculture*, vol. 179, p. 105865, Dec. 2020, doi: 10.1016/j.compag.2020.105865.