



# Implementasi CART-Real Adaboost dalam Memprediksi Minat Pelanggan Membeli Sepatu

Moch. Anjas Aprihartha<sup>1\*</sup>, Jus Prasetya<sup>2</sup>, dan Sefri Imanuel Fallo<sup>3</sup>

<sup>1</sup>Program Studi PJJ Informatika, Universitas Dian Nuswantoro

<sup>2</sup>Alumni Magister Universitas Gadjah Mada

<sup>3</sup>Program Studi Matematika, Universitas San Pedro

\*Correspondence: E-mail: [anjas.aprihartha@dsn.dinus.ac.id](mailto:anjas.aprihartha@dsn.dinus.ac.id)

## ABSTRAK

Pembelajaran mesin merupakan bidang ilmu yang berkaitan pengembangan algoritma komputer untuk mengubah data menjadi suatu tindakan cerdas. Dalam pembelajaran mesin tidak luput dari memahami algoritma pembelajaran mesin. Salah satu algoritma pembelajaran mesin yang populer adalah *supervised learning*. Algoritma *supervised learning* umumnya digunakan dalam memecahkan masalah prediksi. Penelitian ini bertujuan untuk menerapkan algoritma *supervised learning* menggunakan metode *Classification and Regression Tree (CART)* dan *CART-Real Adaptive Boosting* dalam memprediksi minat pelanggan membeli sepatu. Hasil penelitian menunjukkan performa model *CART* menghasilkan akurasi sebesar 77,5% dan *AUC* sebesar 0,711 yang menandakan model cukup baik. Sedangkan performa model *CART-Real Adaboost* diperoleh model terbaik pada kedalaman pohon berada di level 6 atau level 8. Model tersebut menghasilkan akurasi sebesar 85,71% dan *AUC* sebesar 0,8225 yang menandakan model baik. Ini menjadikan *CART-Real Adaboost* menjadi model terbaik dibandingkan model *CART*.

© 2024 Kantor Jurnal dan Publikasi UPI

## ABSTRACT

*Machine learning is a field of science related to the development of computer algorithms to transform data into intelligent actions. In machine learning does not escape understanding machine learning algorithms. One popular machine learning algorithm is supervised learning. Supervised learning algorithms are commonly used in solving prediction problems. This study aims to implement supervised learning algorithms using CART and CART-Real Adaboost methods in predicting customer interest in buying shoes. The results of the study obtained the performance of the CART model resulted in an accuracy of 77.5% and an AUC of 0.711 which indicates that the model is quite good. While the performance of the CART-Real Adaboost model obtained the best model at tree depth level 6 or level 8. The model obtained an accuracy of 85.71% and an AUC of 0.8225 which indicates a good model. This makes CART-Real Adaboost the best model compared to the CART model.*

© 2024 Kantor Jurnal dan Publikasi UPI

## INFORMASI ARTIKEL

### Sejarah Artikel:

Diterima 15 Maret 2024

Direvisi 22 April 2024

Disetujui 25 April 2024

Tersedia online 1 Mei 2024

Dipublikasikan 2 Mei 2024

### Kata Kunci:

*CART,*  
*Prediksi,*  
*Real Adaboost,*  
*Sepatu,*  
*Supervised Learning.*

### Keywords:

*CART,*  
*Prediction,*  
*Real Adaboost,*  
*Shoes,*  
*Supervised Learning.*

## 1. PENDAHULUAN

Manusia selalu dikelilingi oleh data. Banjir data ini telah menyebabkan beberapa orang menyatakan bahwa kita telah memasuki era *big data*. Manusia selalu dikelilingi oleh data. Hal yang membuat era saat ini menjadi unik ketika kita sadar memperoleh data yang mudah. Lebih besar dan lebih banyak kumpulan data yang menarik semakin mudah diakses hanya dengan pencarian di *website*. Saat ini manusia hidup dikelilingi dengan sejumlah besar data yang dapat langsung diproses oleh mesin. Sebagian besar informasi ini memiliki potensi untuk menginformasikan pengambilan keputusan melalui cara yang sistematis dan masuk akal.

Bidang studi yang berkaitan dengan pengembangan algoritma komputer untuk mengubah data menjadi tindakan cerdas atau mengubah data menjadi informasi dikenal sebagai pembelajaran mesin (Putri, et al., 2023). Bidang ini berasal dari lingkungan di saat data yang tersedia, metode statistik, dan daya komputasi berkembang dengan cepat dan bersamaan. Pertumbuhan data diperlukan daya komputasi tambahan, yang pada gilirannya memacu pengembangan statistik metode untuk menganalisis kumpulan data besar. Ini menciptakan siklus kemajuan yang memungkinkan data yang lebih besar dan lebih menarik untuk dikumpulkan.

Dalam pembelajaran mesin tidak luput dari memahami algoritma pembelajaran mesin. Salah satu algoritma pembelajaran mesin yang populer adalah *supervised learning* (Alfarizi, et al, 2023). Algoritma *supervised learning* umumnya digunakan dalam memecahkan masalah prediksi atau klasifikasi. Contoh dari algoritma *supervised learning* antara lain *decision tree*, *naive bayes*, regresi logistik, dll.

Penting untuk mengetahui seberapa akurat model dalam memprediksi data yang telah diberikan. Seringkali model memberikan prediksi yang tidak tepat sehingga mengakibatkan kekacauan dalam penelitian. Tentu saja ini akan menjadi masalah bagi peneliti. Untuk mengatasi hal tersebut maka diperkenalkan algoritma "bantu" yang disebut *Adaboost*.

*Adaboost* kependekan dari "*adaptive boosting*" merupakan algoritma yang paling populer dalam meningkatkan akurasi model. Algoritma AdaBoost adalah algoritma pembelajaran berulang yang melatih dan merakit beberapa pengklasifikasi lemah menjadi pengklasifikasi yang kuat untuk mencapai klasifikasi yang sangat akurat (Hu, et al, 2020).

Pada artikel ini dipaparkan beberapa studi kasus terkait implementasi algoritma *supervised learning*. Penelitian pertama yang menjadi referensi dilakukan oleh Jefi et al. (2019). Ia menerapkan metode *decision tree* tipe C4.5 berbasis PSO dalam memprediksi minat pelanggan dalam pembelian produk sepatu. Hasil penelitian diperoleh ketepatan model dalam memprediksi data secara keseluruhan sebesar 78,16%. Penelitian kedua dalam bidang medis berasal dari Li, et al. (2021) yang mengaplikasikan dua metode sekaligus yaitu CART dan CART+*Adaboost* sebagai perbandingan dalam mendeteksi penyakit hati. Dalam studi kasusnya dipilih model CART+*Adaboost* yang cocok dalam mendeteksi penyakit hati. Penelitian ketiga dalam bidang teknik oleh Liu, et al. (2020), yang memprediksi massa bebatuan untuk tunneling yang efisien dan pencegahan bahaya pada mesin bor terowongan dengan menerapkan metode CART dan CART+*Adaboost*. Berdasarkan hasil penelitiannya diperoleh model CART+*Adaboost* merupakan model yang terbaik.

Artikel ini akan memperbaharui penelitian oleh Jefi et al. (2019) atas saran dari peneliti sebelumnya dengan menambahkan metode *data mining* seperti *Adaboost*. Berdasarkan paparan dari penelitian terdahulu, pada studi kasus kali ini akan mengimplementasikan metode CART dan CART+*Adaboost*. Tujuan penelitian ini untuk mengetahui akurasi model mana yang terbaik dalam memprediksi minat pelanggan dalam pembelian sepatu.

## 2. METODE

### 2.1 Data dan Variabel Penelitian

Data yang diaplikasikan pada penelitian ini bersumber dari Jefi, *et al.*, (2019). Pengumpulan data tersebut berasal dari perusahaan di Purwakarta. Data terdiri dari delapan variabel dengan 490 sampel data. Variabel dependen merupakan status penjualan yang terdiri dari dua katagori yaitu laku dan tidak laku. Sedangkan variabel independen adalah model, warna, ukuran, kelamin, bahan *upper*, merk, dan jenis sepatu dengan setiap variabel independen bertipe katagorik. Pembagian data *training* dan data *testing* menggunakan rasio 80:20. Jumlah data *training* sebanyak 392 sampel data dan jumlah data *testing* sebanyak 98 sampel data.

### 2.2 Classification and Regression Tree (CART)

Metode *Classification and Regression Tree* (CART) merupakan metode nonparametrik yang paling sukses dalam menghadapi masalah klasifikasi atau analisis regresi (Ghiasi, et al., 2017). Algoritma CART secara bebas dapat mempelajari segala bentuk fungsi pemetaan dari sampel data training yang digunakan. Sebagai metode bersifat nonparametrik maka CART tidak bergantung pada distribusi tertentu. Namun membutuhkan lebih banyak data training dalam pengolahan data dibandingkan dengan metode *naive bayes* atau analisis diskriminan linear. Pencilan pada parameter input tidak terlalu memengaruhi kinerja CART (Ghiasi, et al., 2020).

Dalam pembentukan model, CART menggunakan kriteria *index gini* karena dapat diperluas dengan memasukkan kerugian simetris (*symmetrized cost*). Misalkan  $D$  merupakan himpunan data berisi  $n$  sampel dan  $m$  variabel independen dengan setiap sampel  $i$  direpresentasikan sebagai vektor  $(x_1, x_2, x_3, \dots, x_m)$  dan salah satu kelas pada variabel dependen adalah  $c$ . *Index Gini* pada variabel  $j$  didefinisikan sebagai berikut (Bouke, Abdulah &..., 2023).

$$Gini(j) = 1 - \sum (P(c_k|x_j))^2 \quad (1)$$

dengan  $P(c_k|x_j)$  merupakan probabilitas kelas  $c_k$  pada variabel independen ke  $x_j$  yang dihitung dengan membagi jumlah sampel dengan kelas  $c_k$  dengan jumlah total sampel untuk setiap pemisahan. Indeks Gini dihitung untuk setiap fitur dan memberi peringkat berdasarkan kepentingannya.

### 2.3 Real Adaptive Boosting (Real Adaboost)

*Adaboost* merupakan perwakilan dari semua algoritma *boosting*. Algoritma ini dapat membangun pengklasifikasi yang kuat dari pengklasifikasi yang lemah. Algoritma *adaboost* memiliki peran sebagai alat yang dikombinasikan dengan algoritma klasifikasi seperti *decision tree*, regresi logistik, *naive bayes*, dsb. dalam meningkatkan akurasi model tunggal. Pada fase pelatihan data dari *adaboost*, bobot pengklasifikasi dasar dan distribusi probabilitas dari sampel pelatihan dihitung berdasarkan tingkat kesalahan pelatihan (Xing, et al., 2024).

Algoritma *real adaboost* dapat dipahami sebagai pembelajaran statistik yang membangun model regresi logistik aditif melalui pendekatan ekspansi Taylor dalam minimalisasi fungsi kerugian eksponensial. Regresi logistik adalah pendekatan populer yang digunakan dalam

ilmu statistika untuk mengatasi permasalahan klasifikasi. Pada kelas biner, model logistik aditif diformulasikan sebagai berikut (Friedman, et al., 2000).

$$F_M(x) = \ln \frac{P(Y=1|x)}{P(Y=-1|x)} = \sum_{m=1}^M f_m(x) \tag{2}$$

dengan  $M$  merupakan kombinasi linear sebanyak  $M$  dan  $f_m(x)$  merupakan adalah fungsi dari subset kecil dari variabel masukan yang telah ditentukan sebelumnya.

Misalkan minimalisasi kriteria fungsi eksponensial sebagai berikut:

$$J(F) = E(e^{-yF(x)}|x) \tag{3}$$

dengan  $E$  menyatakan ekspektasi populasi atau rata-rata sampel.

Dalam meminimalkan fungsi kerugian eksponensial maka dilakukan iterasi sebanyak  $m$  kali. Setelah mencapai iterasi ke- $m$  diperoleh model yang dinotasikan  $F_m(x) = \sum_{s=1}^m c_s f_s(x)$ . Pada iterasi berikutnya model diperkuat dengan menambahkan fungsi  $c_{m+1}f_{m+1}(x)$ .

$$F_{m+1}(x) = F_m(x) + c_{m+1}f_{m+1}(x) \tag{4}$$

Meminimalkan fungsi kerugian dapat dilakukan dengan dua cara, dengan memperhatikan  $c_{m+1}$  atau  $f_{m+1}$ . Misalkan  $c_{m+1}$  dan  $x$  tetap, maka fungsi kerugian pada  $F_{m+1}(x)$  didefinisikan sebagai berikut:

$$J(F_{m+1}(x)) = J(F_m(x) + c_{m+1}f_{m+1}(x)) \tag{5}$$

Melalui pendekatan ekspansi Taylor *second order* persamaan (5) menjadi

$$J(F_{m+1}(x)) = E \left[ e^{-yF_m(x)} \left( 1 - yc_{m+1}f_{m+1} + \frac{1}{2}y^2c_{m+1}^2f_{m+1}^2 \right) \right] \tag{6}$$

Ketika  $y \in \{1, -1\}$  dan  $f_{m+1} \in \{1, -1\}$  maka  $y^2 = f_{m+1}^2 = 1$ . Sehingga

$$J(F_{m+1}(x)) = E \left[ e^{-yF_m(x)} \left( 1 - yc_{m+1}f_{m+1} + \frac{1}{2}c_{m+1}^2 \right) \right] \tag{7}$$

Perhatikan bahwa  $f_{m+1}$  berapa pada suku kedua sehingga meminimalkan fungsi kerugian pada persamaan (7) sama dengan memaksimalkan suku kedua yang menghasilkan ekspektasi bersyarat. Selain itu, untuk setiap  $c > 0$  maka  $c_{m+1}$  dapat dihilangkan dari fungsi.

$$\max_f E[e^{-yF_m(x)} y f_{m+1}(x) | x] \tag{8}$$

Didefinisikan bobot  $w^{(m)} = w^{(m)}(y, x) = e^{-yF_m(x)}$ . Mengoptimalkan persamaan (8) sama dengan memaksimalkan ekspektasi bobot bersyarat.

$$\max_f E_w[y f_{m+1}(x) | x] \tag{9}$$

dengan  $E_{w^{(m)}}(y|x) = \frac{E(wy|x)}{E(w|x)}$  yang mengacu pada ekspektasi bobot bersyarat.

Perhatikan bahwa,

$$\begin{aligned} E_{w^{(m)}}[yf_{m+1}(x)|x] &= P_{w^{(m)}}(y = 1|x)f_{m+1}(x) - P_{w^{(m)}}(y = -1|x)f_{m+1}(x) \\ &= [P_{w^{(m)}}(y = 1|x) - P_{w^{(m)}}(y = -1|x)]f_{m+1}(x) \\ &= E_w(y|x)f_{m+1}(x) \end{aligned} \quad (10)$$

dengan  $P_{w^{(m)}}(y|x) = \frac{E(w|y,x)P(y|x)}{E(w|x)}$ . Solusi pada  $f_{m+1}(x)$  adalah

$$f_{m+1}(x) = \begin{cases} 1; & E_w(y|x) > 0 \\ -1; & E_w(y|x) \leq 0 \end{cases} \quad (11)$$

Selanjutnya meminimalkan fungsi kerugian (5) terhadap  $c_{m+1}$ .

$$\begin{aligned} J(F_m(x) + c_{m+1}f_{m+1}(x)) &= E(e^{-y(F_m(x) + c_{m+1}f_{m+1}(x))}) \\ &= E(e^{-yF_m(x)} e^{-yc_{m+1}f_{m+1}(x)}) \\ &= E_{w^{(m)}}(e^{-yc_{m+1}f_{m+1}(x)}) \end{aligned} \quad (12)$$

Persamaan (12) dapat dibagi menjadi dua bagian yakni probabilitas terklasifikasi benar dan probabilitas terklasifikasi salah.

$$E_{w^{(m)}}(e^{-yc_{m+1}f_{m+1}(x)}) = P_w(y = f_{m+1}(x))e^{-c_{m+1}} + P_w(y \neq f_{m+1}(x))e^{c_{m+1}} \quad (13)$$

Nilai minimum dari  $c_{m+1}$  dapat dihitung melalui turunan pertama dari  $E_{w^{(m)}}(e^{-yc_{m+1}f_{m+1}(x)})$  terhadap  $c_{m+1}$  diperoleh

$$\begin{aligned} c_{m+1} &= \frac{1}{2} \ln \frac{P_w(y=f_{m+1}(x))}{P_w(y \neq f_{m+1}(x))} \\ &= \frac{1}{2} \ln \frac{1-\varepsilon_{m+1}}{\varepsilon_{m+1}} \end{aligned} \quad (14)$$

Sehingga persamaan (4) menjadi

$$F_{m+1}(x) = F_m(x) + \left(\frac{1}{2} \ln \frac{1-\varepsilon_{m+1}}{\varepsilon_{m+1}}\right) f_{m+1}(x) \quad (15)$$

Karena  $y \in \{1, -1\}$  dan  $f_{m+1} \in \{1, -1\}$  maka  $I$  didefinisikan sebagai berikut:

$$I(y, f_{m+1}) = \begin{cases} 0; & f_{m+1}(x_i) = y_i \\ 1; & f_{m+1}(x_i) \neq y_i \end{cases} \quad (16)$$

Maka,

$$\varepsilon_{m+1} = \frac{\sum_{i=1}^N w_i^{(m+1)} I(y_i, f_{m+1}(x_i))}{\sum_{i=1}^N w_i^{(m+1)}} \quad (17)$$

Selanjutnya memperbaharui bobot vektor  $w_i^{(m+1)}$ . Bobot vektor  $w_i^{(m+1)}$  dinyatakan sebagai berikut:

$$\begin{aligned}
 w^{(m+1)} &= e^{-y_i F_{m+1}(x_i)} \\
 &= e^{-y_i(F_m(x_i)+c_{m+1}f_{m+1}(x_i))} \\
 &= w_i^{(m)} e^{-y_i c_{m+1} f_{m+1}(x_i)} \\
 &= w_i^{(m)} e^{I(y_i, f_{m+1}(x_i)) \frac{1}{2} \ln \frac{1-\epsilon_{m+1}}{\epsilon_{m+1}}} \tag{18}
 \end{aligned}$$

## 2.4 Evaluasi Kinerja Model

Dalam masalah klasifikasi, penting untuk menentukan kinerja model terkait dengan klasifikasi benar atau klasifikasi salah. Evaluasi kinerja model didasarkan pada jumlah pengujian objek data yang diprediksi benar atau salah. Hasil tersebut ditabulasikan dalam *confusion matrix* (Patro & Patra, 2014).

**Tabel 1. Confusion Matrix**

	<b>Kelas Prediksi: Positif</b>	<b>Kelas Prediksi: Negatif</b>
<b>Kelas Observasi: Positif</b>	Positif Benar (PB)	Negatif Salah (NS)
<b>Kelas Observasi: Negatif</b>	Positif Salah (PS)	Negatif Benar (NB)

Pengukuran ketepatan prediksi dapat dilakukan dengan beberapa perhitungan melalui akurasi, sensitifitas, spesifisitas, presisi dan *Area Under the ROC Curve* (AUC).

**Tabel 2. Rumus Ukuran Kinerja Model**

<b>Pengukuran</b>	<b>Rumus</b>
Akurasi	$\frac{PB+NB}{PB+NS+PS+NB}$
Sensitifitas	$\frac{PB}{PB+NS}$
Spesifisitas	$\frac{NB}{PS+NB}$
Presisi	$\frac{PB}{PB+NS}$
AUC	$\frac{1 + \left(\frac{PB}{PB+NS}\right) - \left(1 - \frac{NB}{PS+NB}\right)}{2}$

Panduan dasar untuk mengklasifikasikan keakuratan tes diagnostik menggunakan AUC disajikan sebagai berikut (Nindrea, et al., 2014).

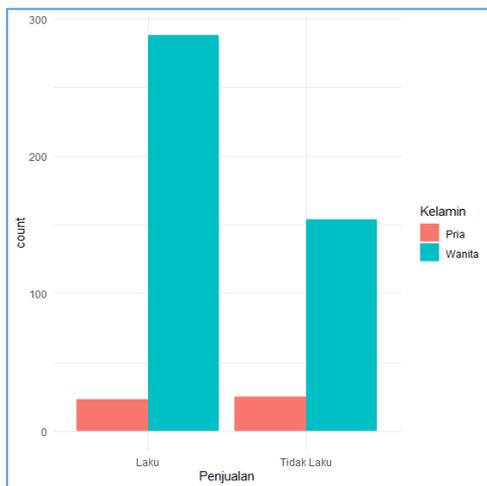
- 0,90 – 1,00 = Klasifikasi Sempurna
- 0,80 – 0,89 = Klasifikasi Baik
- 0,70 – 0,79 = Klasifikasi Cukup Baik
- 0,60 – 0,69 = Klasifikasi Kurang Baik
- 0,50 – 0,59 = Klasifikasi Gagal

## 3. HASIL DAN PEMBAHASAN

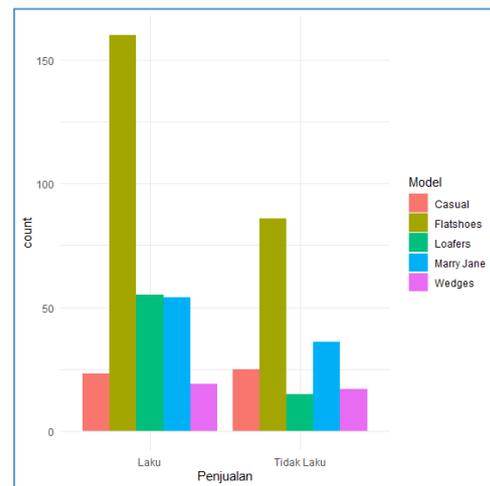
### 3.1 Eksplorasi Data

Gambaran terhadap karakteristik data terkait faktor-faktor yang mempengaruhi penjualan sepatu maka digunakan analisis deskriptif. Dalam penelitian Jefa et al., (2017)

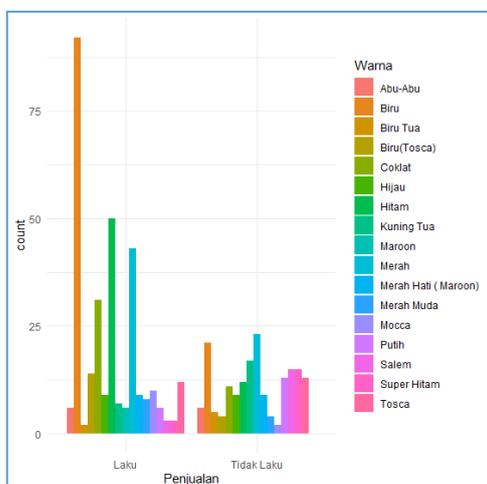
faktor-faktor yang diduga memengaruhi penjualan sepatu adalah model sepatu, warna sepatu, jenis kelamin pembeli, bahan *upper* sepatu, merk sepatu, kategori sepatu, ukuran sepatu. Gambar 1 diketahui sepatu dengan warna biru memiliki peminat besar sedangkan sepatu dengan warna biru tua kurang diminati pembeli. Sementara itu, sepatu warna merah maroon telah habis terjual. Gambar 2 diketahui bahwa model *flatshoes* merupakan model terlaris dengan jumlah penjualan 160 buah sedangkan model *wedges* kurang diminati dengan jumlah penjualan hanya 17 buah. Pada Gambar 3, diketahui pembelian sepatu lebih banyak terjadi pada wanita dibandingkan laki-laki. Hal ini disebabkan lebih banyak pilihan model sepatu bagi kaum wanita. Gambar 4 diketahui sepatu jenis bahan *Canvas 8 Oz* lebih laku dengan jumlah penjualan 61 buah. Berbanding terbalik dengan bahan *Kentucky* hanya terjual 15 buah. Pada Gambar 5, sepatu dengan ukuran 41 lebih banyak dicari dibandingkan sepatu ukuran 44. Selain itu, stok kosong sepatu dengan ukuran 43. Pada Gambar 6 diketahui sepatu merk *St. Yves* paling laku dengan penjualan sebanyak 112 buah kemudian disusul oleh merk *Connexion* dan merk *Nevada*. Pada Gambar 7, sepatu formal lebih laku terjual dibandingkan nonformal dengan penjualan terbesar sebanyak 160 buah.



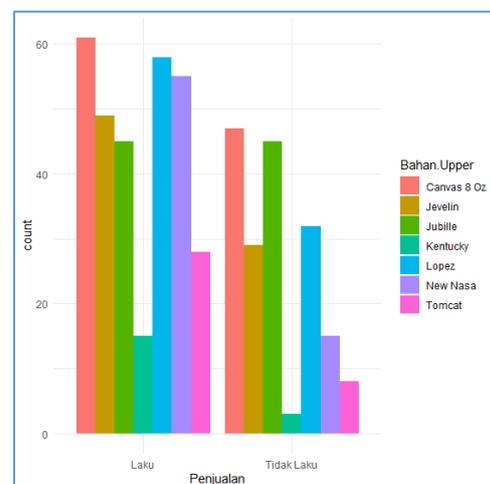
Gambar 1. Penjualan vs Warna



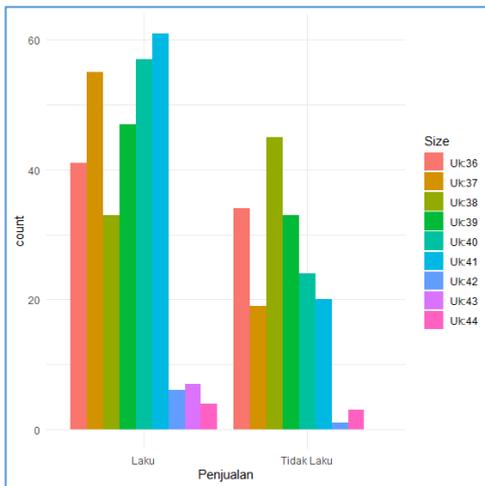
Gambar 2. Penjualan vs Model



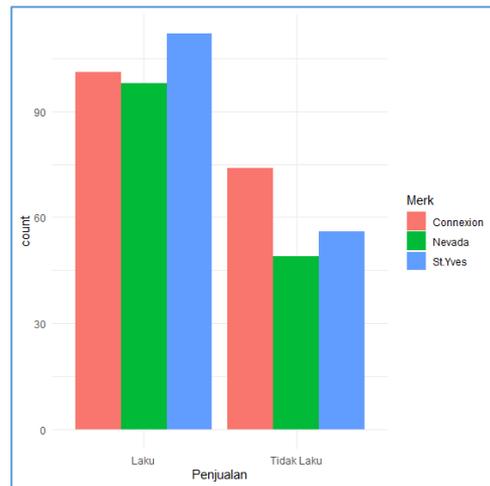
Gambar 3. Penjualan vs Jenis Kelamin



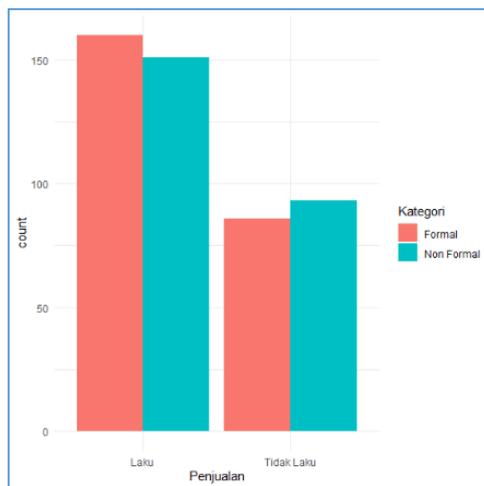
Gambar 4. Penjualan vs Bahan Upper



Gambar 5. Penjualan vs Size



Gambar 6. Penjualan vs Merk



Gambar 7. Penjualan vs Kategori

### 3.2 Hasil Prediksi Model CART

Sebelum mengaplikasikan algoritma *real adaboost*, terlebih dahulu membuat model prediksi dengan algoritma CART menggunakan persamaan (1). Hasil prediksi disajikan dalam *confusion matrix* pada Tabel 3.

Tabel 3. Confusion Matrix CART

	Kelas Prediksi: Laku	Kelas Prediksi: Tidak Laku	Jumlah
Kelas Observasi: Laku	56	12	68
Kelas Observasi: Tidak Laku	12	18	30
Jumlah	68	30	98

Jika dilihat pada Tabel 3 dari total 68 sepatu pada kelas laku, model dapat memprediksi dengan tepat terdapat 56 sepatu berstatus laku dan 12 sepatu adalah kesalahan klasifikasi. Kemudian, dari total 30 sepatu pada kelas tidak laku, model hanya dapat memprediksi dengan

tepat terdapat 18 sepatu berstatus tidak laku dan sisanya adalah kesalahan klasifikasi. Berdasarkan Tabel 2 dapat dihitung performa model CART.

$$Akurasi = \frac{56+18}{98} = 0,775 = 77,5\%$$

$$Sensitifitas = \frac{56}{68} = 0,823 = 82,3\%$$

$$Spesifisitas = \frac{18}{30} = 0,6 = 60\%$$

$$Presisi = \frac{56}{68} = 82,35\%$$

$$AUC = \frac{1+0,823-(1-0,6)}{2} = 0,711$$

Berdasarkan perolehan dari performa model dapat dijelaskan bahwa model berhasil mengidentifikasi penjualan sepatu dengan tepat secara keseluruhan sebesar 77,5% sedangkan 22,5% adalah kesalahan klasifikasi. Model dapat mendeteksi kelas laku dengan benar sebesar 82,3%. Model dapat mendeteksi kelas tidak laku dengan benar sebesar 60%. Model berhasil memprediksi 82,35% dari total prediksi laku dengan benar.

Nilai AUC bertujuan untuk mengukur uji diagnostik. Batasan dalam AUC berada pada rentang 0 sampai 1. Jika AUC mendekati 1 dapat diartikan uji diagnostik semakin baik. Pada model CART diperoleh nilai AUC sebesar 0,711 yang mengindikasikan model cukup baik.

### 3.3 Hasil Prediksi Model CART-Real Adaboost

Penerapan *CART-real adaboost* bertujuan untuk meningkatkan akurasi dari model CART yang diperoleh sebelumnya. Proses awal memberikan bobot yang sama untuk setiap data pada data training kemudian membentuk pengklasifikasi lemah dengan algoritma CART pada persamaan (1). Kemudian, pengklasifikasi lemah  $f_m$  dipanggil untuk mengklasifikasikan sampel, dan tingkat kesalahan klasifikasi yang sesuai  $\varepsilon_m$  dihitung dengan persamaan (17). Pada  $\varepsilon_m$  digunakan untuk memperbarui bobot setiap sampel dengan persamaan (18) dan untuk menghitung bobot  $c_m$  dari pengklasifikasi lemah  $f_m$  diiterasi berikutnya. Proses berulang terjadi berulang-ulang. Akhirnya, pengklasifikasi kuat  $F_M$  dikumpulkan dari pengklasifikasi lemah ini dan bobotnya yang sesuai.

Faktor yang memengaruhi akurasi dari *CART-real adaboost* adalah tingkat kedalaman pohon dan jumlah iterasi. Kedua hiperparameter ini biasanya merupakan sumber *overfitting*. Semakin dalam pohonnya maka semakin kompleks modelnya. Pohon yang memiliki kedalaman terbesar dapat menyesuaikan pelatihan data dengan baik. Untuk mencegah hal tersebut terjadi maka perlu membatasi kedalaman pohon dengan menetapkan nilai *max depth* optimum. *Max depth* merupakan parameter yang mampu mengontrol kedalaman pohon. Sementara itu, jumlah iterasi yang ditetapkan dalam pembentukan model sebanyak 100 iterasi.

Pada Tabel 4 diperlihatkan terdapat sembilan model dengan masing-masing tingkat kedalaman pohon yang berbeda-beda. Hasil uji diperoleh bahwa nilai AUC hampir mengalami peningkatan secara monoton karena kompleksitas model meningkat. Misalkan ketika kedalaman pohon berada pada level 2 hingga 6 maka pengujian AUC juga meningkat dari

0,6166 hingga 0,8225. Nilai *max depth* optimum berada dalam kedalaman level ke-6 dan level ke-8. Ini berarti pohon tidak akan tumbuh lebih dalam dari 6 level atau 8 level. Perolehan hasil performa model mereka adalah sama. Nilai akurasi, spesifisitas, presisi dan AUC paling tinggi diantara model-model lainnya. Pada model CART-*real adaboost* diperoleh nilai AUC sebesar 0,8225 yang mengindikasikan model yang baik. Hasil ini senada dengan penelitian oleh (Chuan, et al., 2021) dengan menerapkan metode *adaboost* dalam studi kasusnya. Ia menjelaskan model dengan kinerja baik pada data *training* berada pada kedalaman pohon level 8 dan iterasi ke-100, AUC yang dihasilkan sebesar 0,63.

**Tabel 4.** Kinerja Model CART- *Real Adaboost*

Model	Max Depth	Akurasi	Sensitifitas	Spesifisitas	Presisi	AUC
1	2	76.53%	100.00%	23.30%	74.72%	0.6166
2	3	77.55%	88.23%	53.33%	81.08%	0.7078
3	4	79.59%	91.17%	53.33%	81.87%	0.7225
4	5	81.83%	89.70%	63.33%	84.72%	0.7651
<b>5</b>	<b>6</b>	<b>85.71%</b>	<b>91.17%</b>	<b>73.33%</b>	<b>88.57%</b>	<b>0.8225</b>
6	7	84.69%	92.64%	66.66%	86.30%	0.7968
<b>7</b>	<b>8</b>	<b>85.71%</b>	<b>91.17%</b>	<b>73.33%</b>	<b>88.57%</b>	<b>0.8225</b>
8	9	84.69%	91.17%	70.00%	87.32%	0.8059
9	10	80.61%	85.29%	70.00%	86.56%	0.7764

Pada penelitian Jefa et al., (2017) diaplikasikan metode C4.5 berbasis *particle swarm optimization* dan memberikan akurasi model lebih baik yaitu 78,16% dibandingkan dengan model C4.5 yaitu 73,88%. Sementara itu, dengan pembagian data yang sama performa akurasi model CART yaitu 77,5% dan performa akurasi model CART-*Real Adaboost* yaitu 85,71%. Terjadi peningkatan sebesar 8,21% antara akurasi model CART dengan model CART-*Real Adaboost*. Ini menjadikan CART-*Real Adaboost* menjadi model terbaik dibandingkan model lainnya.

#### 4. KESIMPULAN

Berdasarkan hasil pengujian metode CART dan metode *real adaboost* yang diaplikasikan dalam CART dapat disimpulkan performa model CART diperoleh akurasi sebesar 77,5%; sensitifitas 82,3%; spesifisitas 60%; presisi sebesar 82,35%; dan AUC sebesar 0,711 yang menandakan model cukup baik. Sedangkan performa model CART dengan *real adaboost* diperoleh model terbaik pada kedalaman pohon berada di level 6 atau level 8. Model menghasilkan akurasi sebesar 85,71%; sensitifitas sebesar 91,17%; spesifisitas sebesar 73,33%; presisi sebesar 88,57%; dan AUC sebesar 0,8225 yang menandakan model baik. Ini menunjukkan model CART-*real adaboost* merupakan model yang layak digunakan dalam memprediksi penjualan sepatu karena memberikan performa model lebih baik dibandingkan model CART.

## 5. DAFTAR PUSTAKA

- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk machine learning dan deep learning. *Karimah Tauhid*, 2(1), 1-6.
- Bouke, M. A., Abdullah, A., Frnda, J., Cengiz, K., & Salah, B. (2023). BukaGini: A stability-aware gini index feature selection algorithm for robust model performance. *IEEE Access*, 11, 59386 – 59396.
- Chuan, Y., Zhao, C., He, Z., & Wu, L. (2021). The success of AdaBoost and its application in portfolio management. *International Journal of Financial Engineering*, 8(02), 2142001.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337-407.
- Ghiasi, M. M., & Mohammadi, A. H. (2017). Application of decision tree learning in modeling CO2 equilibrium absorption in ionic liquids. *Journal of Molecular Liquids*, 242, 594-605.
- Ghiasi, M. M., Zendejboudi, S., & Mohsenipour, A. A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer Methods and Programs in Biomedicine*, 192, 105400.
- Hu, G., Yin, C., Wan, M., Zhang, Y., & Fang, Y. (2020). Recognition of diseased pinus trees in UAV images using deep learning and AdaBoost classifier. *Biosystems Engineering*, 194, 138-151.
- Jefi, J., Puspita, A., & Fahmi, M. (2019). Prediksi peminatan pelanggan dalam penjualan produk sepatu menggunakan metode decision tree berbasis particle swarm optimization pada PT. Baskara Cipta Pratama. *Jurnal Teknik Informatika*, 5(1), 10-17.
- Li, X., Chen, X., & Yuan, Z. (2021, June). Applicable model of liver disease detection based on the improved CART-AdaBoost algorithm. In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 1177-1181). IEEE.
- Liu, Q., Wang, X., Huang, X., & Yin, X. (2020). Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data. *Tunnelling and Underground Space Technology*, 106, 103595.
- Nindrea, R. D., Aryandono, T., Lazuardi, L., & Dwiprahasto, I. (2018). Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. *Asian Pacific Journal of Cancer Prevention*, 19(7), 1747.
- Patro, V. M., & Patra, M. R. (2014). Augmenting weighted average with confusion matrix to enhance classification accuracy. *Transactions on Machine Learning and Artificial Intelligence*, 2(4), 77-91.

Putri, A. I., Syarif, Y., Jayadi, P., Arrazak, F., & Salisah, F. N. (2023). Implementation of decision tree and Support Vector Machine (SVM) algorithm for stunting risk prediction. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 349-357.

Xing, H. J., Liu, W. T., & Wang, X. Z. (2024). Bounded exponential loss function based AdaBoost ensemble of OCSVMs. *Pattern Recognition*, 148, 110191.