

## Comparison of Naïve Bayes, CART, dan CART Adaboost Methods in Predicting Tire Product Sales

### Perbandingan Metode *Naive Bayes*, CART, dan CART dengan *Discrete Adaboost* dalam Memprediksi Penjualan Produk Ban

Moch. Anjas Aprihartha<sup>1)</sup>, Fitri Astutik<sup>2)</sup>, Nani Sulistianingsih<sup>3)</sup>

<sup>1)</sup>Program Studi PJJ Informatika, Fakultas Ilmu Komputer, Dian Nuswantoro University

<sup>2,3)</sup>Program Studi Sistem dan Teknologi Informasi, Fakultas Teknik,  
Universitas Muhammadiyah Mataram

**Email:** [anjas.aprihartha@dsn.dinus.ac.id](mailto:anjas.aprihartha@dsn.dinus.ac.id), [fitri.astutik@ummat.ac.id](mailto:fitri.astutik@ummat.ac.id)<sup>2)</sup>,  
[nani.sulistianingsih@ummat.ac.id](mailto:nani.sulistianingsih@ummat.ac.id)<sup>3)</sup>

Received: 24 February 2024, revised: 24 March 2024, accepted: 26 March 2024

#### Abstract

Data mining is a term to describe the process of moving through large databases in search of certain previously unknown patterns. In finding certain patterns, you need a supporting technique, called machine learning. Machine learning involves learning hidden patterns in data and further using patterns to classify or predict an event related to a problem. One of the problems can be solved with machine learning such as predicting the sales rate of tire products. This can help companies predict tire products that are selling well in the market. In producing an accurate prediction model, it will be compared with decision tree classification methods of CART, CART + Discrete Adaboost, and Naive Bayes applied to tire sales data by PT. Mitra Mekar Mandiri. The results of the study based on successive model performance evaluations are model Naive Bayes < model CART < model CART+Discrete Adaboost. The Discrete Adaboost model with a data proportion of 90:10 is the best model for predicting tire sales. The accuracy, sensitivity and specificity values for the model were 79.17%; 89.47%; and 68.84%. The AUC value is 0.8 which indicates the model is good.

**Keywords:** CART, discrete adaboost, naive bayes.

#### Abstrak

*Data mining* merupakan istilah untuk mendeskripsikan proses perpindahan melalui database besar untuk mencari pola tertentu yang sebelumnya tidak diketahui. Dalam menemukan pola tertentu maka perlu teknik yang mendukung yang disebut *machine learning*. *Machine learning* melibatkan pembelajaran pola tersembunyi dalam data dan selanjutnya menggunakan pola untuk mengklasifikasikan atau memprediksi suatu peristiwa yang terkait dengan masalah. Salah satu masalah yang dapat diselesaikan dengan *machine learning* seperti memprediksi tingkat



penjualan produk ban. Hal ini dapat membantu perusahaan dalam memprediksi produk ban yang laris dipasaran. Dalam menghasilkan model prediksi yang akurat maka akan dibandingkan metode klasifikasi *decision tree* tipe CART, CART+Discrete Adaboost, dan Naive Bayes yang diaplikasikan pada data penjualan ban oleh PT. Mitra Mekar Mandiri. Hasil penelitian menunjukkan bahwa model Naive Bayes < model CART < model CART+Discrete Adaboost berdasarkan hasil evaluasi kinerja model. Model CART+Discrete Adaboost dengan proporsi data 90:10 adalah model terbaik dalam memprediksi penjualan ban. Nilai akurasi, sensitifitas dan spesifisitas untuk model tersebut masing-masing 79,17%; 89,47%; dan 69,84%. Nilai AUC diperoleh 0,8 yang mengindikasikan model yang baik

**Kata kunci:** CART, discrete adaboost, naive bayes.

## 1. PENDAHULUAN

*Data mining* merupakan istilah untuk mendeskripsikan proses perpindahan melalui database besar untuk mencari pola tertentu yang sebelumnya tidak diketahui [14]. Dalam menemukan pola tersebut maka perlu proses penambangan data menggunakan alat dan teknik tertentu untuk menganalisis data dari sumber data yang besar. Teknik yang mendukung dan paling banyak digunakan dalam *data mining* adalah *machine learning*. *Machine learning* melibatkan pembelajaran pola tersembunyi dalam data dan selanjutnya menggunakan pola untuk mengklasifikasikan atau memprediksi suatu peristiwa yang terkait dengan masalah [11].

Berbagai metode klasifikasi yang telah banyak digunakan oleh para peneliti. Terdapat tiga metode klasifikasi yang cukup populer seperti CART, Adaboost, dan Naive Bayes. *Classification and Regression Tree* (CART) merupakan metode klasifikasi nonparametrik yang sukses mengatasi masalah klasifikasi dan regresi [7]. Sebagai metode yang bersifat nonparametrik maka tidak bergantung pada distribusi tertentu. Sebaliknya metode Naive Bayes merupakan metode parametrik sehingga sangat bergantung pada distribusi yang digunakan. Dibandingkan dengan CART, metode Naive Bayes hanya membutuhkan data *training* dalam jumlah kecil pada proses pengolahannya. Adaboost kependekan dari "*adaptive boosting*" merupakan metode gabungan yang paling populer dalam meningkatkan akurasi model. Adaboost bekerja dengan menggabungkan beberapa pengklasifikasi "lemah" menjadi kesatuan pengklasifikasi yang lebih akurat dalam memecahkan masalah nonlinear yang sulit [11].

Penelitian pertama sebagai referensi atau acuan dilakukan oleh Wijaya [18]. Ia menerapkan metode Naive Bayes dalam mengklasifikasikan tingkat penjualan ban. Tujuan dari penelitiannya agar perusahaan dapat menentukan tata letak barang yang baik pada produk terlaris dipasaran sehingga dapat mempercepat sirkulasi. Peneliti menyarankan membandingkan model Naive Bayes dengan metode klasifikasi yang lain untuk mengetahui perbandingan akurasi dan tingkat kesamaan tiap-tiap model. Selain itu, ketersediaan informasi berupa data yang diperlukan sangat membantu dalam melanjutkan penelitian ini.

Penelitian oleh Eliana, dkk. [4] menerapkan metode CART dalam mengetahui karakteristik tingkat penjualan laptop di *e-commerce*. Hasil penelitian diketahui nilai akurasi model diperoleh sebesar 95,10% sehingga model CART layak digunakan dalam mengklasifikasikan tingkat penjualan laptop di di *E-commerce*. Penelitian lain oleh Hafiz [9] menjelaskan pemanfaatan metode CART dalam memprediksi omset sepatu pria. Melalui pengujian dengan data dari PT. Matahari Department Store Thamrin Plaza Mede diperoleh hasil persentase *correctly classified instance* adalah sebesar 100%.

Penelitian oleh Nurlaela [16] menerapkan metode Adaboost dalam memprediksi pendapatan penjualan film. Penelitian ini bertujuan untuk memprediksi film yang memiliki tingkat kesuksesan tinggi. Hasil penelitian menunjukkan dengan metode Adaboost memberikan akurasi 84,44%. Penelitian lain oleh Abdullah, dkk. [1] melakukan analisa penjualan *video game* menggunakan metode Adaboost. Penelitian ini bertujuan agar memberikan kemudahan penjualan

sehingga memperoleh keuntungan maksimal pada pengembang *video game*. Hasil penelitian menunjukkan akurasi model sebesar 93,10%.

Dari ketiga metode yang telah dijelaskan, penelitian ini bertujuan untuk mengetahui performa klasifikasi mana yang paling baik dengan melihat ukuran akurasi, sensitifitas, spesifisitas, dan *Area Under the ROC Curve* (AUC) sehingga akan dipilih satu metode yang layak digunakan dalam memprediksi penjualan ban oleh PT. Mitra Mekar Mandiri.

## 2. METODOLOGI PENELITIAN

### 2.1 Data dan Variabel Penelitian

Data pada penelitian ini diperoleh dari tugas akhir Wijaya [18] yang menerapkan metode klasifikasi *Naive Bayes* dalam studi kasus tersebut. Data yang dimiliki diperoleh dari data bulanan penjualan ban tahun Januari 2013 hingga Desember 2017 pada PT. Mitra Mekar Mandiri. Rata-rata penjualan produk ban tahun Januari 2013 hingga Desember 2017 sebanyak 195 unit. Hasil ini yang akan menjadi acuan dalam mengkategorikan penjualan produk ban.

Total data terdiri dari 2400 data yang akan dibagi menjadi *data training* dan *data testing*. Variabel independen adalah tipe, pola, ukuran sedangkan variabel dependen adalah penjualan.

**Tabel 2.1.** Variabel Penelitian

Variabel	Kategori
Penjualan	Penjualan dibagi menjadi dua kategori yaitu: <ol style="list-style-type: none"> <li>1. Penjualan &lt; 195 unit</li> <li>2. Penjualan <math>\geq</math> 195 unit</li> </ol>
Tipe	Tipe dibagi menjadi dua kategori yaitu: <ol style="list-style-type: none"> <li>1. Tubeless</li> <li>2. Tube Type</li> </ol>
Pola	Pola dibagi menjadi lima kategori yaitu: <ol style="list-style-type: none"> <li>1. Cityforce</li> <li>2. Dozer</li> <li>3. Flemmino</li> <li>4. Raven</li> <li>5. Spartax</li> </ol>
Ukuran	Ukuran dibagi menjadi empat kategori yaitu: <ol style="list-style-type: none"> <li>1. 80/90-14</li> <li>2. 90/90-14</li> <li>3. 80/90-17</li> <li>4. 90/90-17</li> </ol>

### 2.2 Decision Tree Tipe CART

*Decision tree* merupakan metode klasifikasi yang sangat sering digunakan dalam berbagai kasus. Algoritma *Classification and Regression Tree* (CART) salah satu algoritma dalam membangun model *decision tree* yang secara rekursif mempartisi data berdasarkan variabel dengan nilai *gini index* tertinggi [3]. *Gini index* adalah kriteria berbasis ketidakmurnian dalam mengukur perbedaan antara distribusi probabilitas dari variabel dependen [14]. *Gini index* mudah diimplementasikan pada algoritma CART terutama pada variabel yang berdimensi tinggi. Selain itu, *gini index* digunakan untuk memilih variabel-variabel independen yang layak digunakan di masing-masing simpul internal pada model *decision tree*. Diberikan kumpulan data  $S$ , *gini index* didefinisikan sebagai berikut [11].

$$Gini(S) = 1 - \sum_{i=0}^{c-1} p_i^2 \quad (2.1)$$

dengan  $c$  adalah jumlah kelas yang ditentukan,  $p_i = S_i/S$  adalah frekuensi relatif kelas  $C_i$  dalam himpunan,  $s_i$  adalah jumlah sampel pada kelas  $C_i$ , dan  $C_i$  adalah kelas untuk  $i = 1, 2, \dots, c - 1$ .

Pemisahan (*split*) pada variabel menjadi himpunan bagian  $k$  dari  $S_i$  dihitung sebagai jumlah pembobot *gini index* dari himpunan bagian yang dihasilkan.

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i) \quad (2.2)$$

dengan  $n_i$  adalah jumlah sampel dalam *subset*  $S_i$  setelah pemisahan,  $n$  adalah jumlah total sampel dalam simpul yang diberikan.

### 2.3 Discrete Adaptive Boosting

*Boosting* merupakan teknik tambahan yang diterapkan dalam upaya meningkatkan model prediksi yang dihasilkan dari metode utama seperti metode CART. Teknik ini yang dapat mengubah beberapa model yang "lemah" menjadi satu model yang "kuat" [15]. *Boosting* bekerja dengan membangun banyak model secara berurutan dalam proses berulang menggunakan hasil evaluasi dari model sebelumnya.

*Discrete adaptive boosting* atau *discrete adaboost* merupakan teknik turunan dari *boosting*. Teknik ini dapat secara efektif memprediksi kelas biner pada data yang berdimensi tinggi [5].

*Discrete adaboost* pada dasarnya adalah  $m$  buah kombinasi linear yang dinyatakan sebagai berikut [6].

$$C_m(x) = \sum_{j=1}^m a_m k_m(x) \quad (2.3)$$

dengan  $C_m(x)$  adalah model pohon akhir,  $k_m(x) = y \in \{-1, 1\}$  adalah model tunggal yang baru dibangun, dan  $a_m$  adalah tingkat pembelajaran.

Misalkan  $C_m(x) = 1/2 \sum_{j=1}^m a_m k_m(x)$  dengan  $1/2$  merupakan konstanta. *Discrete adaboost* menggunakan fungsi kerugian eksponensial yang didefinisikan sebagai berikut:

$$l(y_i, \hat{y}_i) = e^{-y_i \hat{y}_i} = e^{-y_i C_m(x)} \quad (2.4)$$

Oleh karena itu, total kesalahan pada  $N$  sampel dinyatakan sebagai berikut:

$$E = \sum_{i=1}^N w_i^{(m)} e^{-\frac{1}{2} y_i a_m k_m(x_i)} \quad (2.5)$$

dengan  $w_i^{(m)} = e^{-\frac{1}{2} y_i C_{m-1}(x_i)}$ .

Persamaan (2.5) dibagi menjadi dua bagian yakni data terklasifikasi benar dan data terklasifikasi salah. Kriteria data yang terklasifikasi benar apabila berada pada interval 0,5 sampai 1 sedangkan data yang terklasifikasi salah berada pada interval 0 sampai 0,499.

$$E = \sum_{y_i = k_m(x_i)} w_i^{(m)} e^{-\frac{1}{2} a_m} + \sum_{y_i \neq k_m(x_i)} w_i^{(m)} e^{\frac{1}{2} a_m} \\ E = W_c e^{-\frac{1}{2} a_m} + W_e e^{\frac{1}{2} a_m} \quad (2.6)$$

Nilai minimum dari  $a_m$  dapat dihitung melalui turunan pertama dari  $E$  terhadap  $a_m$

$$a_m = \ln \frac{W_c}{W_e} = \ln \frac{1 - W_e}{W_e} = \ln \frac{1 - \varepsilon_m}{\varepsilon_m} \quad (2.7)$$

dengan  $a_m$  tingkat pembelajaran ke  $m$ . Karena  $y_i \in \{-1, 1\}$  dan  $k_m(x_i) \in \{-1, 1\}$  maka  $I$  didefinisikan sebagai berikut:

$$I(k_m(x_i), y_i) = \begin{cases} 0; & k_m(x_i) = y_i \\ 1; & k_m(x_i) \neq y_i \end{cases} \quad (2.8)$$

Jadi

$$\varepsilon_m = \frac{\sum_{i=1}^N w_i^{(m)} I(k_m(x_i), y_i)}{\sum_{i=1}^N w_i^{(m)}} \quad (2.9)$$

*Classifier*  $C_m(x)$  dengan  $a_m$  pada (2.7) menjadi

$$\begin{aligned} C_m(x) &= C_{m-1}(x) + a_m k_m(x) \\ C_m(x) &= C_{m-1}(x) + \left( \ln \frac{1-\varepsilon_m}{\varepsilon_m} \right) k_m(x) \end{aligned} \quad (2.10)$$

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

$$\begin{aligned} w_i^{(m)} &= e^{-y_i C_m(x)} \\ &= e^{-y_i (C_{m-1}(x) + (\ln \frac{1-\varepsilon_m}{\varepsilon_m}) k_m(x))} \\ w_i^{(m)} &= w_i^{(m-1)} e^{I(k_m(x_i), y_i) \ln \frac{1-\varepsilon_m}{\varepsilon_m}} \end{aligned} \quad (2.12)$$

## 2.4 Naive Bayes

*Naive Bayes* merupakan metode yang diciptakan oleh Thomas Bayes dengan menerapkan probabilitas bersyarat sebagai dasar pembentukan model. Model *Naive Bayes* bekerja dengan cara mengkalkulasi probabilitas kelas pada variabel dependen dari masing-masing variabel yang ada. Misalkan  $Y$  adalah variabel dependen dan  $X$  mewakili variabel independen. Estimasi dari  $\Pr(Y = C_l|X)$  dinyatakan sebagai probabilitas hasil pada kelas ke- $l$  dapat dihitung dengan aturan Bayes sebagai berikut [12].

$$\Pr(Y = C_l|X) = \frac{\Pr(Y=C_l)\Pr(X|Y=C_l)}{\Pr(X)} \quad (2.13)$$

dengan  $\Pr(Y = C_l|X)$  biasa disebut sebagai probabilitas posterior kelas,  $\Pr(Y = C_l)$  adalah probabilitas prior *outcome*,  $\Pr(X)$  adalah probabilitas dari variabel independen, dan  $\Pr(X|Y = C_l)$  adalah probabilitas  $X$  dengan syarat  $Y = C_l$  telah terjadi, dan  $C_l$  adalah kelas ke- $l$ .

Untuk menghitung probabilitas bersyarat  $\Pr(X|Y = C_l)$  dapat menggunakan fungsi probabilitas kepadatan untuk setiap anggota pada variabel independen,

$$\begin{aligned} \Pr(X|Y = C_l) &= \prod_{j=1}^p \Pr(X_j|Y = C_l) \\ &= \Pr(X_1|Y = C_l) \times \Pr(X_2|Y = C_l) \times \dots \times \Pr(X_p|Y = C_l) \end{aligned} \quad (2.14)$$

## 2.5 Evaluasi Kinerja Model

Dalam menguji ketepatan model prediksi terdapat beberapa cara yang umum digunakan melalui pengukuran akurasi, sensitifitas, dan spesifisitas [10].

**Tabel 2.2.** Pengukuran Kinerja Model

Pengukuran	Formula
Akurasi	$\frac{TP+TN}{P+N}$
Sensitifitas	$\frac{TP}{P}$
Spesifisitas	$\frac{TN}{N}$

dengan *true positive* (TP) adalah jumlah data positif yang diprediksi dengan benar, *true negative* (TN) adalah jumlah data negatif yang diprediksi dengan benar, *false positive* (FP) adalah jumlah data positif yang diprediksi tidak benar, dan *false negative* (FN) adalah jumlah data negatif yang diprediksi secara tidak benar. P adalah jumlah TP dan FN dan N adalah jumlah FP dan TN. Istilah-istilah ini dapat disusun ke dalam *confusion matrix*.

Tabel 2.3. *Confusion Matrix*

	Kelas Prediksi: Ya	Kelas Prediksi: Tidak	Jumlah
Kelas Observasi: Ya	$TP$	$FN$	$P$
Kelas Observasi: Tidak	$FP$	$TN$	$N$
Jumlah	$P'$	$N'$	$P + N$

Selain menggunakan ketiga pengukuran diatas, terdapat ukuran yang disebut *Area Under the ROC Curve* (AUC). Pengukuran ini dapat dihitung sebagai kurva ROC yang dinyatakan sebagai berikut [13].

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2}$$

Ukuran yang menjadi dasar penilaian dari hasil klasifikasi menggunakan AUC [8].

- 0,90 – 1,00 = Klasifikasi Baik Sekali
- 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 *Data Training* dan *Data Testing*

Data yang digunakan dalam analisis ini berjumlah 2400 data dengan tiga variabel independen yaitu tipe, pola, dan ukuran sedangkan variabel dependen adalah penjualan. Variabel penjualan bersifat biner dengan dua kategori yaitu tidak laris dengan penjualan dibawah 195 unit dan laris dengan penjualan diatas 195 unit. Langkah selanjutnya membagi *data training* dan *data testing* dalam beberapa proporsi.

Tabel 3.1. Pembagian *Data Training*

No	Proporsi Data	Jumlah <i>Data Training</i>	Jumlah <i>Data Testing</i>
1	60 : 40	1440	960
2	70 : 30	1680	720
3	80 : 20	1920	480
4	90 : 10	2160	240

Proses pembentukan model klasifikasi dengan software RGui 4.3.2. Dalam membangun model CART menggunakan fungsi *rpart*, model CART dengan *Discrete Adaboost* menggunakan fungsi *ada*, sedangkan model *Naive Bayes* menggunakan fungsi *naiveBayes*. Setelah model terbentuk langkah selanjutnya melatih model dengan *data training*. Hasil dari melatih model dapat diketahui dari evaluasi kinerja model pada *data training*. Model klasifikasi akan merubah kemampuan kerja pada dirinya untuk menyesuaikan dengan data yang diberikan saat latihan sehingga model tersebut dapat memahami informasi pada *data training* [4]. Hasil pemahaman model dalam menggali informasi *data training* dapat dievaluasi dalam kinerja model melalui perhitungan akurasi, sensitifitas, dan spesifisitas.

#### 3.2 Perbandingan Hasil Prediksi Model Dengan *Data Training*

Pada bagian ini disajikan hasil prediksi model CART dalam tabel *confusion matrix* dengan proporsi *data training* 70% dan *data testing* 30%.

## JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Moch. Anjas Aprihartha, Fitri Astutik, Nani Sulistianingsih

**Tabel 3.2.** *Confusion Matrix* Model CART

	Kelas Prediksi: 1	Kelas Prediksi: 2	Jumlah
Kelas Observasi: 1	726	213	939
Kelas Observasi: 2	168	573	741
Jumlah	894	786	1680

Berdasarkan Tabel dijelaskan bahwa kelas 1 (Penjualan < 195 unit) yang tepat terprediksi sebagai kelas 1 sebanyak 726 data sedangkan terjadi *missclassified* antara kelas 2 (Penjualan  $\geq$  195 unit) yang terprediksi di kelas 1 sebanyak 168 data. Kelas 2 yang tepat terprediksi sebagai kelas 2 sebanyak 573 data. *Missclassified* terjadi antara kelas 1 yang terprediksi di kelas 2 sebanyak 213 data. Langkah selanjutnya menghitung akurasi, sensitifitas, spesifisitas, dan AUC berdasarkan *confusion matrix* pada Tabel

$$\begin{aligned} \text{Akurasi} &= \frac{726+573}{1680} = \frac{1299}{1680} = 0,7732 = 77,32\% \\ \text{Sensitifitas} &= \frac{726}{939} = 0,7731 = 77,31\% \\ \text{Spesifisitas} &= \frac{573}{741} = 0,7732 = 77,32\% \\ \text{AUC} &= \frac{1+0,7731-(1-0,7732)}{2} = 0,7735 \end{aligned}$$

Hasil evaluasi kinerja setiap model pada masing-masing proporsi data yang berbeda disajikan pada Tabel 3.3.

**Tabel 3.3.** Kinerja Model Pada *Data Training*

Proporsi Data	Pengukuran	CART	<i>Discrete Adaboost</i>	<i>Naive bayes</i>
60 : 40	Akurasi	77,57%	78,06%	73,33%
	Sensitifitas	90,66%	86,64%	83,92%
	Spesifisitas	62,48%	68,16%	61,14%
	AUC	0,7657	0,7740	0,7252
70 : 30	Akurasi	77,32%	77,56%	72,44%
	Sensitifitas	77,31%	86,35%	81,43%
	Spesifisitas	77,32%	67,56%	62,21%
	AUC	0,7735	0,7695	0,7182
80 : 20	Akurasi	76,82%	77,55%	72,40%
	Sensitifitas	76,45%	84,14%	78,92%
	Spesifisitas	77,24%	70,17%	65,08%
	AUC	0,7684	0,7715	0,7199
90 : 10	Akurasi	76,20%	77,27%	74,03%
	Sensitifitas	84,67%	85,81%	82,47%
	Spesifisitas	66,83%	67,80%	64,68%
	AUC	0,7574	0,7680	0,7357
<b>Rata-rata AUC</b>		0,7662	0,771	0,7247

Berdasarkan Tabel 3.3 dapat dijelaskan bahwa ketiga model klasifikasi yang dilatih dengan *data training* memiliki AUC pada rentang 0,70 hingga 0,79 artinya model cukup baik dalam

## JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Moch. Anjas Aprihartha, Fitri Astutik, Nani Sulistianingsih

memahami informasi dari *data training*. Urutan model dalam memahami informasi data dinilai berdasarkan rata-rata AUC adalah model *Naive Bayes* < model CART < model CART+*Discrete Adaboost*.

Setelah proses training dilakukan, tahap selanjutnya adalah melakukan evaluasi terhadap performa model yang sudah dilatih sebelumnya. Dalam prosesnya performa model akan diuji menggunakan *data testing*. Proses ini menunjukkan kemampuan kerja model dalam memprediksi dengan benar pada data baru yang belum pernah dilihat sebelumnya. Kemampuan generalisasi model klasifikasi yang dipelajari pada sampel *data training* terbatas pada data baru menjadi tantangan yang signifikan [2]. Semakin tinggi generalisasi dari sebuah model maka akan semakin meningkatkan akurasi prediksi.

### 3.3 Perbandingan Hasil Prediksi Model Dengan *Data Testing*

Hasil evaluasi kinerja setiap model pada masing-masing proporsi data pada *data testing* disajikan pada Tabel 3.4.

**Tabel 3.4.** Kinerja Model Pada *Data Testing*

Proporsi Data	Pengukuran	CART	<i>Discrete Adaboost</i>	<i>Naive Bayes</i>
60 : 40	Akurasi	75,21%	76,56%	69,48%
	Sensitifitas	88,91%	85,98%	81,38%
	Spesifisitas	61,62%	67,22%	57,68%
	AUC	0,7526	0,7660	0,6952
70 : 30	Akurasi	77,22%	78,47%	70,56%
	Sensitifitas	80,56%	81,47%	78,59%
	Spesifisitas	73,97%	75,07%	62,47%
	AUC	0,7726	0,7936	0,7066
80 : 20	Akurasi	77,50%	77,50%	70,62%
	Sensitifitas	73,50%	73,50%	76,07%
	Spesifisitas	81,30%	81,30%	65,45%
	AUC	0,7740	0,7740	0,7075
90 : 10	Akurasi	77,08%	79,17%	70,82%
	Sensitifitas	86,84%	89,47%	81,85%
	Spesifisitas	68,25%	69,84%	61,11%
	AUC	0,7754	0,800	0,7134
<b>Rata-rata AUC</b>		0,7686	0,7834	0,7056

Penelitian Wijaya [18] menggunakan tiga kategori sebagai variabel dependen. Selain itu, peneliti menggunakan keseluruhan data sebagai data *training*. Pada penelitian ini variabel dependen disederhanakan menjadi dua katagori yaitu penjualan  $\geq 195$  unit dan  $< 195$  unit. Hasil uji model *Naive Bayes* memperoleh akurasi yang berbeda-beda dikarenakan perbedaan pembagian data *training* dan data *testing*. Pada Tabel 3.4, model *Naive Bayes* terbaik ketika rasio data 90:10 dengan akurasi model sebesar 70,82%.

Model *Naive Bayes* pada proporsi data 60 : 40 memiliki AUC sebesar 0,6952 sehingga model tersebut termasuk dalam katagori model kurang baik. Model CART dan model *Discrete Adaboost* memiliki kinerja model yang sama pada akurasi, sensitifitas, spesifisitas, dan AUC. Hasil kinerja

model berturut-turut berdasarkan rata-rata AUC adalah model *Naive Bayes* < model CART < model CART+*Discrete Adaboost*.

Hasil yang senada didapatkan pada penelitian Nurlalela [16] dimana eksperimen dengan model tunggal seperti *Naive Bayes* menghasilkan akurasi 83,22%. Setelah dilakukan *boosting* dengan algoritma *Adaboost* terjadi peningkatan akurasi menjadi 84,24%. Penelitian oleh Abdullah [1] dimana perbandingan kedua model *Adaboost* dengan regresi linear memiliki hasil lebih baik dibandingkan model tunggal regresi linear.

Pada model model CART+*Discrete Adaboost*, persentase spesifisitas untuk proporsi data 60 : 40 tergolong rendah dibandingkan model-model CART+*Discrete Adaboost* lainnya. Misalkan model model CART+*Discrete Adaboost* memiliki spesifisitas sebesar 67,22% yang berarti bahwa model berhasil mengenali 67,22% penjualan  $\geq 195$  unit secara akurat. Sisanya sebesar 32,78% berarti model masih melewatkan penjualan  $\geq 195$  unit yang sebenarnya.

Selain itu, model model CART+*Discrete Adaboost* dengan proporsi data 90:10 paling baik karena memiliki akurasi sebesar 79,17%. Ini berarti ketepatan model berhasil mengenali secara akurat penjualan  $\geq 195$  unit dan < 195 unit sebesar 79,17%. Akurasi ini paling tinggi diantara model-model klasifikasi lainnya. Sehingga pada kasus ini dipilih model model CART+*Discrete Adaboost* dengan proporsi *data training* 90:10 sebagai model yang cocok dalam memprediksi penjualan ban oleh PT. Mitra Mekar Mandiri.

#### 4. KESIMPULAN

Berdasarkan hasil pembahasan dapat disimpulkan bahwa pada proses melatih *data training* ketiga model klasifikasi memiliki AUC pada rentang 0,70 hingga 0,79 artinya model cukup baik dalam memahami informasi dari *data training*. Hasil penelitian berdasarkan evaluasi kinerja model pada *data training* maupun *data testing* berturut-turut adalah model *Naive Bayes* < model CART < model CART+*Discrete Adaboost*. Model CART+*Discrete Adaboost* dengan proporsi data 90:10 adalah model terbaik dalam memprediksi penjualan ban oleh PT. Mitra Mekar Mandiri. Nilai akurasi 79,17% yang berarti ketepatan model memprediksi secara akurat penjualan ban  $\geq 195$  unit dan penjualan < 195 unit sebesar 79,17%, nilai sensitifitas 89,47% berarti ketepatan model memprediksi penjualan < 195 unit sebesar 89,47%, dan sensitifitas 69,84% berarti ketepatan model memprediksi penjualan  $\geq 195$  unit sebesar 69,84%. Nilai AUC diperoleh 0,8 yang mengindikasikan model yang baik.

#### DAFTAR PUSTAKA

- [1]. Abdullah, H. A., Putra, D. R. D., & Azhar, Y, 2022. Analisa Penjualan Video Game Menggunakan Metode Ensemble. *Just IT: Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, Vol. 12, No. 3, 8-16.
- [2]. Alhajeri, M. S., Alnajdi, A., Abdullah, F., & Christofides, P. D, 2023. On Generalization Error of Neural Network Models and Its Application to Predictive Control of Nonlinear Processes. *Chemical Engineering Research and Design*, Vol. 189, 664-679.
- [3]. Bouke, M. A., Abdullah, A., ALshatebi, S. H., Abdullah, M. T., & El Atigh, H, 2023. An Intelligent Ddos Attack Detection Tree-Based Model Using Gini Index Feature Selection Method. *Microprocessors and Microsystems*, Vol. 98, 104823.
- [4]. Erliani, N., Suryowati, K., & Jatipaningrum, M. T. 2023, Klasifikasi Tingkat Penjualan Laptop Di E-Commerce Menggunakan Algoritma Classification and Regression Tree (CART). *Jurnal Statistika Industri dan Komputasi*, Vol. 8, No. 2, 40-47.

- [5]. Freund, Y., Schapire, R.E. 1997, A Decision Theoretic Generalization of Online Learning and an Application to Boosting. *Journal of Computer and System Sciences*, Vol. 55, 119 - 139.
- [6]. Friedman, J., Hastie, T., & Tibshirani, R., 2000. Additive Logistic Regression a Statistical View of Boosting. *Annals of Statistics*, Vol. 28, No. 2, pp. 337-374.
- [7]. Ghiasi, M. M., & Mohammadi, A. H., 2017. Application of decision tree learning in modelling CO2 equilibrium absorption in ionic liquids. *Journal of Molecular Liquids*, Vol. 242, pp. 594-605.
- [8]. Gorunescu, F, 2011. *Data Mining: Concepts, Models and Techniques*. Springer Science & Business Media, Berlin.
- [9]. Hafiz, M. I, 2019. Pemanfaatan Metode Cart Untuk Memprediksi Omset Sepatu Pria. *Pelita Informatika: Informasi dan Informatika*, Vol. 8, No. 2, 227-235.
- [10]. Han, J., Pei, J., & Tong, H, 2022. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, United States.
- [11]. Kantardzic, M., 2011. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- [12]. Kuhn, M., & Johnson, K, 2013. *Applied predictive modeling*. Springer, New York.
- [13]. López, V., Fernández, A., García, S., Palade, V., & Herrera, F, 2013. An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics. *Information Sciences*, Vol. 250, 113–141.
- [14]. Maimon, O. Z., & Rokach, L, 2014. *Data Mining with Decision Trees: Theory and Applications*. World Scientific, Singapore.
- [15]. Munshi, T. A., Jahan, L. N., Howladar, M. F., & Hashan, M, 2024. Prediction Of Gross Calorific Value from Coal Analysis Using Decision Tree-Based Bagging and Boosting Techniques. *Heliyon*, Vol. 10, No. 1.
- [16]. Nurlaela, D. 2020. Penerapan Adaboost Untuk Meningkatkan Akurasi Naive Bayes Pada Prediksi Pendapatan Penjualan Film. *Inti Nusa Mandiri*, Vol. 14, No. 2, 181-188.
- [17]. Saputra, M. J., & Herdiansyah, M. I, 2022. Penerapan *Naive bayes* Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. *Jurnal Mantik*, Vol. 6, No. 2, 2502-2507.
- [18]. Wijaya, F, 2018. Implementasi Algoritma *Naive Bayes* dalam Klasifikasi Produk Ban Terlaris pada PT. Mitra Mekar Mandiri, *Skripsi*. Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung.