

# ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA NAÏVE BAYES UNTUK MENGLASIFIKASIKAN REKOMENDASI PENJUALAN ONLINE RETAIL

Moh. Irfan Rusdi<sup>1</sup>, Triawan Adi Cahyanto<sup>2</sup>, Reni Umilasari<sup>3</sup>

1410651180

Program Studi Teknik Informatika  
Universitas Muhammadiyah Jember

## Abstrak

*Online retail* adalah suatu bentuk perdagangan elektronik yang memungkinkan konsumen untuk langsung membeli barang atau jasa dari seorang penjual melalui internet tanpa jasa perantara. Kemudahan yang disediakan oleh *Online Retail* mempengaruhi pola masyarakat dalam penggunaan toko online untuk pembelian produk. Maka perlu adanya analisa data untuk mengklasifikasikan data rekomendasi penjualan produk baju menggunakan algoritma *Decicion Tree* dan *Naïve Bayes* untuk membandingkan akurasi dan presisi dimana menjadi tolak ukur dalam penelitian ini. Untuk mendapatkan nilai akurasi dan presisi perlu menggunakan teknik uji *Cross Validation* dimana teknik uji ini membagi data training dan data testing yang bertujuan untuk meminimalisir nilai akurasi dari faktor kebetulan. Setelah dilakukan pengujian dengan teknik *Cross fold Validation* dengan jumlah *K-fold* 2, 5 dan 10 dari algoritma C4.5 dan Naïve Bayes sama-sama mendapat nilai tertinggi pada *10-fold Validation* metode uji data imbang pada skenario ke-8. Algoritma C4.5 mendapat nilai akurasi sebesar 72,00% dan presisi 75,00% sedangkan algoritma Naïve Bayes mendapat nilai akurasi sebesar 72,00% dan presisi 84,62%. Sehingga dapat disimpulkan bahwa kedua algoritma sama-sama sangat akurat dan konsisten dalam penentuan rekomendasi penjualan produk *Online Retail*.

*Kata Kunci: Online Retail, C4.5, Naïve Bayes, K-fold Cross Validation.*

## 1. Pendahuluan

Perkembangan teknologi zaman sekarang mengakibatkan banyak terjadi perubahan di berbagai bidang. Salah satunya terjadi perubahan yang cukup signifikan dibidang industri perdagangan. Pada bidang perdagangan di era modern saat ini banyak terjadi perkembangan pesat dimana bukan terjadi pada pasar tradisional saja melainkan menggunakan jasa online yang sangat mempermudah pelanggan dalam pembelian produk untuk keperluan pribadi maupun keperluan keluarga.

Jasa online juga mengalami perkembangan terus-menerus seiring dengan perkembangan

teknologi. Ada beberapa jenis situs online yang banyak digemari oleh masyarakat dunia seperti *marketplace*, *e-commerce* dan *online retail*. *Online retail* adalah suatu bentuk perdagangan elektronik yang memungkinkan konsumen untuk langsung membeli barang atau jasa dari seorang penjual melalui internet tanpa jasa perantara. Keunggulan dari penggunaan online retail adalah konsumen hanya perlu duduk dan menggunakan *smartphone*-nya untuk mencari produk atau melihat-lihat barang yang akan dibeli dan memesan barang hingga dikirim oleh penjual.

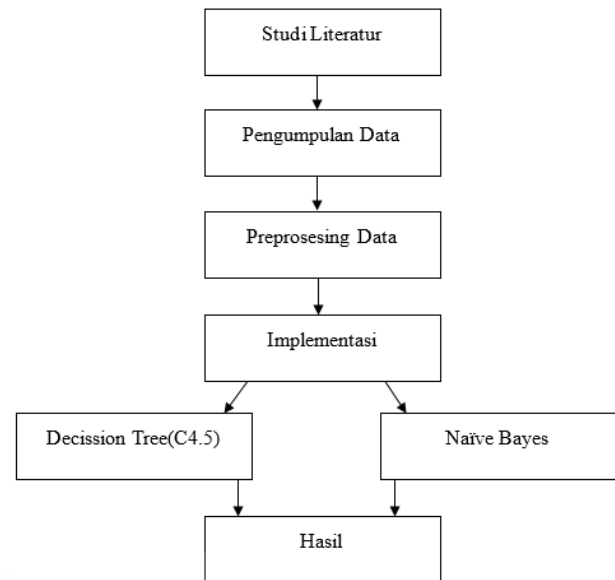
Penggunaan layanan internet sekarang ini menjadi kebutuhan wajib setiap masyarakat. Ketergantungan ini berdampak

pada minat masyarakat untuk menggunakan *Online Retail* dalam kegiatan sehari-hari. Namun ada beberapa faktor yang menjadi perhitungan masyarakat untuk memilih jasa *Online Retail*. Faktor-faktor tersebut diantaranya masyarakat dapat melihat barang terlebih dahulu untuk membelinya, proses transaksi hanya butuh *gadget*, pembelian barang cukup dirumah. Berbagai kemudahan tersebut sangat memanjakan pelanggan untuk membeli produk menggunakan jasa *Online Retail* dan mempengaruhi pola masyarakat dalam penggunaan *Online Retail* untuk pembelian produk.

Berdasarkan uraian diatas maka penulis melakukan analisa dari kumpulan data penjualan produk pada *Online Retail* dengan mengambil data dari situs <https://archive.ics.uci.edu/ml/machine-learning-databases/00289/> untuk klasifikasi rekomendasi produk baju pada *Online Retail* berdasarkan atributnya untuk menghasilkan perbandingan klasifikasi akurasi dan presisi dari algoritma C4.5 (*Decision Tree*) dan algoritma *Naïve Bayes* untuk diproses dalam *Data Mining*.

## 2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan pada Gambar 1, meliputi Studi literatur, Pengumpulan data, Preprosesing data, Implementasi dan Hasil.



Gambar 1. Alur Penelitian

## 1. Studi Literatur

Dalam tahap ini peneliti mempelajari tentang semua data dan informasi yang berkaitan dengan algoritma *Decision Tree* C4.5 dan algoritma *Naïve Bayes* serta semua materi yang berhubungan dengan masalah yang akan dibahas. Dalam penelitian ini referensi diambil dari berbagai sumber, seperti buku, jurnal, e-book, laporan penelitian, situs internet serta sumber ilmiah lainnya yang dinilai dapat menambah wawasan untuk penelitian ini.

## 2. Pengumpulan Data

Data diperoleh dari (UCI) Machine Learning University, yang merupakan data penjualan rekomendasi produk baju. Seperti yang kita ketahui UCI merupakan webside penyedia data internasional yang digunakan dalam berbagai penelitian. Dalam studi kasus Online Retail peneliti mengambil kumpulan data penjualan produk baju dari tahun 2012-2013 dari info yang didapat data ini merupakan kumpulan atau rekap dari [www.aliexpress.com/item/](http://www.aliexpress.com/item/) pada tiap transaksi yang ada, keseluruhan total data ini berjumlah 500 record data dan 14 atribut.

## 3. Preprosesing Data

Data yang peneliti gunakan dalam penelitian ini merupakan dataset penjualan produk baju yang disediakan oleh (UCI) *Machine Learning University*, diunduh dari situs webside [archive.ics.uci.edu](http://archive.ics.uci.edu). Data yang diperoleh sebanyak 300 *record* data yang sudah di *Cleaning* untuk menghapus data dan atribut data yang tidak digunakan atau bernilai *null*, *noisy*, *inconsisten* untuk menghasilkan data yang benar-benar terpakai.

## 4. Implementasi

Pada tahap ini peneliti akan menerapkan data yang telah melalui proses preprocessing data kedalam *RapidMiner Studio* dengan menerapkan metode algoritma C4.5 dan algoritma *Naïve Bayes*.

### 4.1 Perhitungan C4.5

Algoritma C4.5 merupakan ekstensi dari algoritma ID3 dan menggunakan prinsip *Decision Tree* yang mirip (Harriyanto, 2017). Algoritma C4.5 mempunyai kelebihan dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain.

Langkah dalam pemrosesan data menggunakan algoritma C4.5 ada 4 tahapan, yaitu mencari nilai *entropy*, mencari nilai *Gain*, pembentukan pohon keputusan dan *rule*.

#### 1. Entropy

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2(p_i)$$

Dimana :

S : Himpunan kasus

n : Jumlah partisi S

$p_i$  : Jumlah kasus pada partisis ke-i

#### 2. Mencari nilai *Gain*

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n (|S_i|/|S| * Entropy(S_i))$$

Dimana :

S : Himpunan kasus

n : Jumlah partisi atribut A

$|S_i|$  : Jumlah kasus pada kasus ke-i

$|S|$  : Jumlah kasus dalam S

#### 3. Pembuatan pohon keputusan

4. Membuat role berdasarkan pohon keputusan skenario uji keberapa didapat nilai terbaik.
5. Diketahui model klasifikasinya

#### 4.2 Perhitungan Naïve Bayes

*Naïve Bayes* adalah algoritma klasifikasi yang mudah, sederhana dan sering digunakan untuk klasifikasi dokumen. Selain itu *Naïve Bayes* juga digunakan untuk pengambilan keputusan berdasarkan hasil klasifikasi yang diperoleh.

Dalam pemrosesan kedalam algoritma *Naïve Bayes* ada beberapa tahapan, yaitu:

1. Penentuan nilai Prior atau nilai probabilitas
2. Menghitung nilai probabilitas tiap kelas

$$P(H|E_1, E_2, E_3) = \frac{P(E_1|H) * P(E_2|H) * P(E_3|H) * P(H)}{P(E_1) * P(E_2) * P(E_3)}$$

3. Membandingkan hasil kelas PC<sub>1</sub> dan PC<sub>2</sub>
4. Diketahui model klasifikasinya.

### 5. Hasil dan Pembahasan

#### 5.1. Pengujian Algoritma C4.5

Pada tahap ini peneliti akan menerapkan data yang telah melalui proses preprocessing data kedalam *RapidMiner Studio* dengan menerapkan metode algoritma C4.5 dan algoritma *Naïve Bayes*. Pembagian data training dan data testing akan dilakukan dengan menggunakan proses *K-fold Cross Validation*,

Tahap ini setelah peneliti menguji algoritma C4.5 pada metode pengujian data imbang, data tidak imbang 1 dan data tidak imbang 2 pada masing-masing nilai K-fold 2, 5 dan 10. Hasil pengujian tersebut digunakan untuk menganalisis nilai-nilai *K-fold* pada

**Tabel 5.1** Hasil pengujian *Cross Validation* pada algoritma C4.5

	Akurasi	Presisi
K=2	50,88%	47,20%
K=5	57,20%	54,13%
K=10	53,20%	49,53%

Metode uji yang digunakan yaitu Dataimbang sebanyak 125 data YES dan 125 data No, kemudian metode ke 2 Data tidakimbang 1 terdiri dari 150 data Yes dan 100 data No, metode ke 3 yaitu data tidakimbang 2 terdiri dari 150 data No dan 100 data Yes. Berikut pemaparannya:

1. Dataimbang

✘ Pengujian dengan *2-fold Validation*

**Tabel 5.2** Confusion Matrix skenario ke-1

True:	Yes	No
Yes:	36	26
No:	26	37

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%$$

$$\text{Akurasi} = \frac{36 + 37}{37 + 26 + 26 + 37} \times 100\% = 58,40\%$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100\%$$

$$\text{Presisi} = \frac{36}{26 + 37} \times 100\% = 58,06\%$$

**Tabel 5.3** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	58,40%	58,06%
2	52,80%	53,03%

✘ Pada pengujian *5-fold Validation*.

**Tabel 5.4** Confusion Matrix pada skenario ke-1

True:	Yes	No
Yes:	14	10
No:	11	15

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%$$

$$\text{Akurasi} = \frac{14 + 15}{14 + 11 + 10 + 15} \times 100\% = 58,00\%$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100\%$$

$$\text{Presisi} = \frac{14}{11 + 14} \times 100\% = 56,00\%$$

**Tabel 5.5** Hasil perhitungan per-skenario

Data uji ke-	Akurasi	Presisi
1	58,00%	56,00%
2	60,00%	52,00%
3	56,00%	56,00%
4	48,00%	48,00%
5	60,00%	72,00%

✘ Pengujian *10-fold Validation*

**Tabel 5.6** Confusion Matrix skenario uji ke-1

True:	Yes	No
Yes:	9	6
No:	3	7

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%$$

$$\text{Akurasi} = \frac{9 + 7}{9 + 3 + 6 + 7} \times 100\% = 64,00\%$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100\%$$

$$\text{Presisi} = \frac{9}{3 + 7} \times 100\% = 60,00\%$$

**Tabel 5.7** Hasil perhitungan per-skenario

Data uji ke-	Akurasi	Presisi
1	64,00%	60,00%
2	44,00%	41,67%
3	48,00%	46,15%
4	60,00%	56,25%
5	48,00%	46,67%
6	64,00%	61,11%
7	56,00%	62,50%
8	72,00%	75,00%
9	48,00%	50,00%
10	56,00%	56,25%

## 2. Data tidakimbang 1

## ✘ Perhitungan 2-fold Validation

**Tabel 5.8** Confusion Matrix skenario uji ke-1

True:	No	Yes
No:	15	24
Yes:	35	51

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{51+15}{51+24+35+15} \times 100\% = 52,80\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{51}{35+51} \times 100\% = 59,30\%$$

**Tabel 5.9** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	52,80%	59,30%
2	52,00%	60,27%

## ✘ Perhitungan 5-fold Validation

**Tabel 5.10** Confusion Matrix skenario ke-1

True:	No	Yes
No:	6	12
Yes:	14	18

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{18+6}{18+12+14+6} \times 100\% = 48,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{18}{12+18} \times 100\% = 60,00\%$$

**Tabel 5.11** Hasil perhitungan per-skenario

Data uji ke-	Akurasi	Presisi
1	48,00%	60,00%
2	48,00%	63,33%
3	56,00%	66,67%
4	50,00%	63,33%
5	66,00%	66,67%

## ✘ Pengujian pada 10-fold Validation

**Tabel 5.12** Confusion Matrix skenario ke-1

True:	No	Yes
No:	3	6
Yes:	7	9

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{9+3}{9+7+6+3} \times 100\% = 48,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{9}{7+9} \times 100\% = 56,25\%$$

**Tabel 5.13** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	48,00%	56,25%
2	60,00%	72,73%
3	68,00%	76,92%
4	56,00%	62,50%
5	52,00%	60,00%
6	40,00%	50,00%
7	44,00%	52,63%
8	48,00%	57,14%
9	52,00%	58,82%
10	64,00%	68,75%

## 3. Data tidakimbang 2

## ✘ Pengujian pada 2-fold Validation

**Tabel 5.14** Confusion Matrix

True:	No	Yes
No:	54	25
Yes:	21	25

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{25+54}{25+21+25+54} \times 100\% = 63,20\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{25}{21+25} \times 100\% = 54,35\%$$

**Tabel 5.15** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	63,20%	54,35%

2	63,20%	54,55%
---	--------	--------

6	68,00%	60,00%
7	56,00%	42,86%
8	56,00%	42,86%
9	52,00%	40,00%
10	72,00%	63,64%

✘ Pada pengujian *5-fold Validation*

**Tabel 5.16** Confusion Matrix skenario ke-1

True:	No	Yes
No:	22	10
Yes:	8	10

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%$$

$$\text{Akurasi} = \frac{10 + 22}{10 + 10 + 8 + 22} \times 100\% = 64,00\%$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100\%$$

$$\text{Presisi} = \frac{10}{10 + 10} \times 100\% = 50,00\%$$

**Tabel 5.17** Hasil perhitungan per-skenario

Data uji ke-	Akurasi	Presisi
1	64,00%	50,00%
2	58,00%	35,00%
3	50,00%	35,00%
4	60,00%	35,00%
5	46,00%	45,00%

✘ Pada pengujian *10-fold Validation*

**Tabel 5.18** Confusion Matrix skenario ke-1

True:	No	Yes
No:	13	5
Yes:	2	5

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%$$

$$\text{Akurasi} = \frac{5 + 13}{5 + 2 + 5 + 13} \times 100\% = 72,00\%$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100\%$$

$$\text{Presisi} = \frac{5}{2 + 5} \times 100\% = 71,43\%$$

**Tabel 5.19** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	72,00%	71,43%
2	60,00%	50,00%
3	64,00%	55,56%
4	56,00%	42,86%
5	60,00%	50,00%

## 5.2. Pengujian Algoritma Naïve Bayes

Hasil yang diperoleh dari aplikasi RapidMiner Studio untuk algoritma C4.5 dan Naïve Bayes dapat dilihat Tabel 1, proses tersebut menggunakan data sebanyak 300 record data dan di uji dengan metode K-fold Cross Validation.

Tahap ini peneliti menguji algoritma C4.5 dan mendapatkan hasil yang sudah diketahui pada proses pengujian di langkah sebelumnya. Pada pengujian tersebut algoritma C4.5 mendapat hasil tertinggi pada *5-fold Validation* dan akan diuji pada 3 skenario uji.

**Tabel 5.20** Hasil pengujian *Cross Validation* pada algoritma *Naïve Bayes*

	Akurasi	Presisi
K=2	61,20%	59,26%
K=5	63,20%	62,50%
K=10	63,20%	62,77%

Metode uji yang digunakan yaitu Dataimbang sebanyak 125 data YES dan 125 data No, kemudian metode ke 2 Data tidakimbang 1 terdiri dari 150 data Yes dan 100 data No, metode ke 3 yaitu data tidakimbang 2 terdiri dari 150 data No dan 100 data Yes. Berikut pemaparannya:

1. Dataimbang

✕ Pada pengujian *2-fold Validation*

**Tabel 5.21** Confusion Matrix skenario ke-1

True:	Yes	No
Yes:	35	24
No:	27	39

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$= \frac{35+39}{35+27+24+39} \times 100\% = 59,20\%$$

$$\text{Akurasi} = \frac{35+27+24+39}{35+27+24+39} \times 100\% = 59,20\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{35}{27+35} \times 100\% = 56,45\%$$

**Tabel 5.22** Hasil pengujian per-skenario

Uji data ke-	Akurasi	Presisi
1	59,20%	56,45%

2	60,80%	65,08%
---	--------	--------

✕ Pada pengujian *5-fold Validation*.

**Tabel 5.23** Confusion Matrix skenario ke-1

True:	Yes	No
Yes:	12	9
No:	13	16

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$= \frac{12+16}{12+13+9+16} \times 100\% = 56,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{12}{13+12} \times 100\% = 48,00\%$$

**Tabel 5.24** Hasil perhitungan per-skenario

Uji data ke-	Akurasi	Presisi
1	56,00%	48,00%
2	62,00%	60,00%
3	70,00%	80,00%
4	54,00%	44,00%
5	52,00%	40,00%

✕ Pada pengujian *10-fold Validation*

**Tabel 5.25** Confusion Matrix skenario ke-1

True:	Yes	No
Yes:	8	4
No:	4	9

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$= \frac{8+9}{8+4+4+9} \times 100\% = 68,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{8}{4+8} \times 100\% = 66,67\%$$

**Tabel 5.26** Hasil pengujian per-skenario

Uji data ke-	Akurasi	Presisi
1	68,00%	66,67%
2	68,00%	83,33%
3	68,00%	75,00%
4	60,00%	41,67%



5	32,00%	25,00%
6	60,00%	46,15%
7	72,00%	53,85%
8	72,00%	84,62%
9	44,00%	30,77%
10	60,00%	38,46%

2. Data tidak imbang 1

✘ Pengujian 2-fold Validation

**5.27** Confusion Matrix skenario ke-1

True:	No	Yes
No:	20	25
Yes:	30	50

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{50+20}{50+30+25+50} \times 100\% = 56,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{50}{30+50} \times 100\% = 66,67\%$$

**Tabel 5.28** Hasil pengujian per-skenario

Data Uji ke-	Akurasi	Presisi
1	56,00%	66,67%
2	49,60%	49,33%

✘ Pengujian pada 5-fold Validation

**Tabel 5.29** Confusion Matrix skenario ke-1

True:	No	Yes
No:	9	14
Yes:	11	16

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{16+9}{16+14+11+9} \times 100\% = 50,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{16}{14+16} \times 100\% = 53,33\%$$

**Tabel 5.30** Hasil perhitungan per-skenario

Data Uji ke-	Akurasi	Presisi
1	50,00%	53,33%
2	48,00%	53,33%

3	60,00%	70,00%
4	60,00%	70,00%
5	56,00%	56,67%

✘ Pengujian 10-fold Validation

**Tabel 5.31** Confusion Matrix skenario ke-1

True:	No	Yes
No:	4	5
Yes:	6	10

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{10+4}{10+6+5+4} \times 100\% = 56,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{10}{6+10} \times 100\% = 66,67\%$$

**Tabel 5.32** Hasil pengujian per-skenario

Data Uji ke-	Akurasi	Presisi
1	56,00%	66,67%
2	60,00%	60,00%
3	60,00%	73,33%
4	52,00%	66,67%
5	68,00%	73,33%
6	52,00%	60,00%
7	56,00%	53,33%
8	56,00%	66,67%
9	68,00%	73,33%
10	48,00%	40,00%

3. Data tidak imbang 2

✘ Pengujian pada 2-fold Validation

**Tabel 5.33** Confusion Matrix skenario ke-1

True:	No	Yes
No:	49	30
Yes:	26	20

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{10+4}{10+6+5+4} \times 100\% = 56,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{10}{6+10} \times 100\% = 66,67\%$$

**Tabel 5.34** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
1	55,20%	40,00%
2	53,60%	46,00%

1	60,00%	40,00%
2	68,00%	50,00%
3	64,00%	50,00%
4	64,00%	50,00%
5	72,00%	70,00%

✘ Pengujian pada *5-fold Validation*

**Tabel 5.35** Confusion Matrix skenario ke-1

True:	No	Yes
No:	17	9
Yes:	13	11

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{11+17}{11+9+13+17} \times 100\% = 56,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{11}{9+11} \times 100\% = 55,00\%$$

**Tabel 5.36** Hasil perhitungan per-skenario

Data uji ke-	Akurasi	Presisi
1	56,00%	55,00%
2	58,00%	40,00%
3	56,00%	40,00%
4	64,00%	40,00%
5	62,00%	55,00%

✘ Pada pengujian *10-fold Validation*

**Tabel 5.37** Confusion Matrix skenario ke-1

True:	No	Yes
No:	11	6
Yes:	4	4

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

$$\text{Akurasi} = \frac{4+11}{4+6+4+11} \times 100\% = 60,00\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\text{Presisi} = \frac{4}{6+4} \times 100\% = 40,00\%$$

**Tabel 5.38** Hasil pengujian per-skenario

Data uji ke-	Akurasi	Presisi
--------------	---------	---------

6	52,00%	40,00%
7	48,00%	30,00%
8	56,00%	40,00%
9	60,00%	20,00%
10	60,00%	30,00%

Langkah-langkah tiap pengujian dilakukan oleh peneliti hingga mendapat hasil dari setiap pengujian. Pada tabel Confusion Matrix ini berfungsi sebagai tolak ukur untuk menghitung jumlah data yang terklasifikasi dalam proses pengujian oleh aplikasi RapidMiner Studio, tabel Confusion Matrix sangat penting karena digunakan untuk mencari nilai akurasi dan presisi. Selanjutnya pada tabel hasil dari masing-masing pengujian dimana setiap pengujian sesuai jumlah nilai K-fold sehingga dari hasil tersebut peneliti mengambil jumlah nilai tertinggi dari semua pengujian yang sudah dilakukan. Tahap berikutnya merupakan analisis.

### 5.3. Analisis

Setelah melakukan pengujian dan mendapatkan hasil pada algoritma C4.5 dan algoritma *Naïve Bayes*, maka peneliti melakukan pengelompokan hasil tertinggi pada setiap scenario uji dari ketiga metode uji yaitu dataimbang, data tidakimbang 1 dan data tidakimbang 2, yang sudah dilakukan pada algoritma C4.5 dan algoritma *Naïve Bayes*. Hal ini bertujuan untuk mengetahui algoritma terbaik untuk diterapkan pada data produk baju.

**Tabel 5.39** Hasil tertinggi pada setiap scenario uji C4.5

C4.5				
Metode	K fold	Data uji ke-	Akurasi	Presisi

Dataimbang	K=2	1	58,40%	58,06%
	K=5	5	60,00%	72,00%
	K=10	8	72,00%	75,00%
Data tidakimbang 1	K=2	1	52,80%	59,30%
	K=5	5	66,00%	66,67%
	K=10	3	68,00%	76,92%
Data tidakimbang 2	K=2	2	63,20%	54,55%
	K=5	1	64,00%	50,00%
	K=10	1	72,00%	71,43%

**Tabel 5.40** Hasil tertinggi pada setiap scenario uji *Naïve Bayes*

Naïve Bayes				
Metode	K fold	Data uji ke-	Akurasi	Presisi
Dataimbang	K=2	2	60,80%	65,08%
	K=5	3	70,00%	80,00%
	K=10	8	72,00%	84,62%
Data tidakimbang 1	K=2	1	56,00%	66,67%
	K=5	3	60,00%	70,00%
	K=10	5	68,00%	73,33%
Data tidakimbang 2	K=2	1	55,20%	40,00%
	K=5	4	64,00%	40,00%
	K=10	6	72,00%	70,00%

Pada hasil tabel diatas merupakan hasil tertinggi dari masing-masing pengujian namun untuk mengetahui akurasi dan presisi terbaik perlu penyempitan pengelompokan dari hasil-hasil tersebut hingga mencapai hasil akhir yang dapat membandingkan dari algoritma C4.5 dan *Naïve Bayes*.

**Tabel 5.41** Hasil tertinggi C4.5

C4.5				
Metode	K fold	Data uji ke-	Akurasi	Presisi

Data imbang	K=10	8	72,00%	75,00%
Data tidak imbang 1	K=10	3	68,00%	76,92%
Data tidak imbang 2	K=10	1	72,00%	71,43%

**Tabel 5.42** Hasil tertinggi Naïve Bayes

Naïve Bayes				
Metode	K fold	Data uji ke-	Akurasi	Presisi
Data imbang	K=10	8	72,00%	84,62%
Data tidak imbang 1	K=10	5	68,00%	73,33%
Data tidak imbang 2	K=10	6	72,00%	70,00%

**Tabel 5.43** Perbandingan hasil tertinggi pada algoritma C4.5 dan Naïve Bayes

Algoritma	Metode	K fold	Data uji ke-	Akurasi	Presisi
C4.5	Data imbang	K=10	8	72,00%	75,00%
Naïve Bayes	Data imbang	K=10	8	72,00%	84,62%

Pada Tabel 4.41 diketahui bahwa kedua algoritma memiliki hasil akurasi yang sama namun terdapat selisih nilai pada presisi. Namun dalam hal ini peneliti bertujuan untuk mengetahui nilai akurasi dan membandingkannya, sehingga algoritma C4.5 maupun algoritma Naïve Bayes sama-sama akurat dalam penentuan rekomendasi penjualan produk baju. Disisi lain scenario uji sama-sama terletak pada data uji ke-8 dan jumlah K fold sebesar *10-fold Validation* dimana teknik uji dengan data imbang.

## 6. Kesimpulan

Setelah beberapa langkah pengujian yang peneliti lakukan, maka peneliti dapat menarik kesimpulan dari hasil yang sudah didapatkan. Kesimpulan tersebut antara lain:

1. Dengan semua pengujian yang dilakukan oleh peneliti menggunakan aplikasi *RapidMiner Studio* pada teknik uji *Cross Validation* dengan sejumlah metode pengujian pada algoritma C4.5 dan algoritma Naïve Bayes mendapat nilai tertinggi pada *10-fold Validation*, dimana pada *10-fold Validation* pembagian data training berjumlah 90% dengan data testing yang mencapai 10% dari data yang teruji yaitu data imbang yang terdiri dari 125 data bernilai YES dan 125 data bernilai No. Sehingga nilai akurasi maupun presisi pada *10-fold Validation* berpeluang menjadi hasil tertinggi dari semua pengujian yang ada.
2. Berdasarkan seluruh dari hasil skenario pengujian yang sudah dilakukan oleh peneliti didapatkan hasil pada algoritma C4.5 dan Naïve Bayes sama-sama memiliki nilai akurasi dan presisi tertinggi pada scenario uji ke-8 dengan data yang di uji adalah data imbang. Algoritma C4.5 nilai akurasi sebesar 72,00% dan presisi sebesar 75,00% sedangkan algoritma Naïve Bayes nilai akurasi sebesar 72,00% dan presisi sebesar 84,62%.

Dari hasil semua pengujian peneliti lakukan kemudian membandingkan nilai akurasi dan didapatkan hasil yang bernilai sama sehingga dapat ditarik kesimpulan bahwa algoritma C4.5 dan Naïve Bayes sama-sama akurat, namun pada segi nilai presisi algoritma Naïve Bayes lebih unggul dalam penentuan rekomendasi penjualan produk baju *Online Retail*.

## DAFTAR PUSTAKA

- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 dan Naive Bayes untuk prediksi penyakit diabetes. *Jurnal Techno Nusa Mandiri* Vol. XIII, No. 1.
- Fitrihanah. Arifin, M. F., Devi. (2017). Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada, *Jurnal Telekomunikasi dan Komputer*, vol.8, no.2, 2018
- Kusrini Dan Emha. (2009). *Algoritma Data Mining*. Yogyakarta: C.V Andi.
- Larose. (2014). *Discovering Knowledge In Data*. New Jersey: John Wiley & Sons.
- Muflikhah, L., Ratnawati, D. E., Putri, R. R. M. (2018). *Data Mining*. Malang: Ub Press.

- Muljono, Artanti, D. P., et. al. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes. Konferensi Nasional Sistem Informasi, STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018.
- Pattipeilohy, W. F., Wibowo, A., Utari, D. R. (2017). Pemodelan Dan Prototipe Sistem Informasi Untuk Prediksi Pembaharuan Polis Asuransi Mobil Menggunakan Algoritma C.45. Prosiding SNATIF. 783.
- Prasetyo, E. (2012). Data Mining Konsep dan Aplikasi menggunakan MATLAB. Yogyakarta: C.V Andi OFFSET.
- Rahmawati, E. (2015). Analisa Komparasi Algoritma Naive Bayes Dan C4.5 Untuk Prediksi Penyakit Liver. Jurnal Techno Nusa Mandiri, Vol. XII No. 2, September 2015.
- Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus : Kspps / Bmt Al-Fadhila). Magister Teknik Informatika Institut Informatika dan Bisnis Darmajaya, Vol. 02 No. 01 Mei 2016.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. Jurnal Pilar Nusa Mandiri. Volume 13 No.1, Maret 2017.
- Sulianta, F., Dominikus, J. (2010). Data Mining Meramalkan Bisnis Perusahaan. Jakarta: PT. Elex Media Komputindo.
- Susanto, E. D. (2015). Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes. Konferensi Nasional Sistem & Informatika. STMIK STIKOM Bali, 9 – 10 Oktober 2015.
- Widiyanto, I., Prasilowati. S. L. (2015). Perilaku Pembelian Melalui Internet, Jurnal Manajemen Dan Kewirausahaan, Vol.17, No. 2, September 2015.
- Wardani, N. W., Dantes, G. R., Indrawan, G. (2018). Prediksi Customer Churn Dengan Algoritma Decision Tree C4.5 Berdasarkan Segmentasi Pelanggan Pada Perusahaan Retail. Jurnal Resistor, Vol. 1 No 1 – April 2018.