

**PENERAPAN ALGORITMA NAÏVE BAYES UNTUK MENENTUKAN  
KLASIFIKASI PRODUK TERLARIS PADA PENJUALAN PULSA**

***THE APPLICATION OF THE NAÏVE BAYES ALGORITHM TO DETERMINE THE  
CLASSIFICATION OF BEST SELLING PRODUCTS ON PULSES SALES***

Ismasari Nawangsih<sup>1</sup>, Asti Setyaningsih<sup>2</sup>

Program Studi Teknik Informatika Fakultas Teknik, Universitas Pelita Bangsa.

Jl. Inpeksi Kalimalang Tegal Danas, Cikarang

<sup>1</sup>ismasari.n@pelitabangsa.ac.id, <sup>2</sup>astisetyaningsih2@gmail.com

**ABSTRAKSI**

Penelitian ini di latar belakang oleh kemajuan perkembangan teknologi komunikasi dan informasi yang sangat pesat dan semakin murah sehingga membuat masyarakat untuk pulsa telepon selular menjadi kebutuhan yang wajib pada masa kini. Dari bermacam produk pulsa yang ada pada konter RA cell pulsa Telkomsel, pulsa XL, pulsa Indosat dan pulsa 3 penulis mengklasifikasi menjadi laris dan tidak laris. Tujuannya dapat mengetahui penerapan data mining dengan menggunakan algoritma naïve bayes dalam menentukan klasifikasi produk terlaris dan hasil akurasi data terhadap stok penjualan pulsa. Dengan mengumpulkan data sebanyak 600 data menjadi 480 data training dan 120 data testing. Data mining merupakan bentuk penggalian data dalam mengklasifikasi pada sebuah data yang jumlahnya banyak, menggunakan tool Rapid Miner dan algoritma naïve bayes merupakan metode klasifikasi yang banyak digunakan karena sederhana dan akurasi yang tinggi dalam klasifikasi data. Berdasarkan hasil penelitian yang telah di lakukan, jenis produk terlaris pada penjualan pulsa menurut nama produk adalah pulsa Telkomsel. Tingkat akurasi klasifikasi dengan metode naïve bayes menghasilkan nilai accuracy sebesar 97,50%, nilai precision 100 % dan nilai recall 93,48% sehingga metode naïve bayes merupakan metode yang cukup baik dalam penelitian ini.

**Kata kunci :** Pulsa, Klasifikasi, Data Mining, Naïve Bayes.

**ABSTRACT**

*This research is motivated by progress of the development of communication technology and information is very fast and increasingly cheap so that makes the community for mobile phone pulses became a mandatory requirements at the present time. Of various pulse products available at the counter RA Cell Pulses Telkomsel, Pulses XL, pulses Indosat and pulses 3 the author classify as best seller and non-sellers. The goal is to find out the implementation of data mining using the naïve bayes algorithm in determining the classification of best-selling products and the result of the accuracy of the data in the sales of pulses. By collecting 600 data into 480 training data and 120 testing data. Data mining is a form of extracting data in classifying a large amount of data, using the rapid miner application and the naïve bayes algorithm is a classification method that is widely used because of its simple and high accuracy classifying data. Based on the result of research that has been done, the type of product that is most restricted on the sale of pulses by product name is Telkomsel pulses. The level of classification accuracy with the naïve bayes method produces an accuracy value of 97,50 %, a precision value of 100 %, and recall value 93, 48% so the naïve bayes method is good in this research.*

**Keywords:** Pulses, Classification, Data Mining, Naïve Bayes.

**1. PENDAHULUAN**

Kemajuan teknologi saat ini disegala bidang meningkat dengan begitu sangat cepat, kemajuan ini membawa peningkatan kebutuhan komunikasi yang semakin tinggi. Perkembangan alat komunikasi telepon selular yang semakin murah dan pesat membuat kebutuhan masyarakat untuk pulsa elektrik semakin meningkat.

Data mining merupakan bentuk penggalian data yang digunakan untuk menggali pengetahuan dari jumlah data yang besar. Salah satu pendekatan yang dapat

digunakan untuk menganalisis data penjualan pulsa adalah dengan melakukan pengolahan data untuk mengetahui hasil klasifikasi dari data tersebut. Algoritma Metode *Naive Bayes* dapat membantu dalam menentukan produk yang laris dan tidak laris pada penjualan pulsa yaitu dapat menghasilkan data yang memiliki tingkat akurasi yang dapat dimanfaatkan untuk memprediksi kelas dari kasus yang baru. Dengan menggunakan *tool Rapid Miner* dalam proses perhitungannya sehingga dapat diketahui tingkat akurasi data.

Penelitian ini dibuat untuk membantu dalam mengklasifikasi produk yang laris dijual

dalam penjualan pulsa sehingga pemilik konter dapat menentukan stok produk yang banyak diminati masyarakat agar setiap kali ada pembeli barang selalu ada untuk dijual.

**Data Mining**

*Data Mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data [1].

**Klasifikasi**

Klasifikasi adalah suatu cara untuk menganalisis data yang diekstrak menggunakan model untuk menggambarkan kelas data dari data *instance* berdasarkan atribut yang dimiliki oleh data tersebut. Adapun atribut-atribut dapat berupa data categorial, ordinal, integer-value, dan real-value [2].

**Naive Bayes**

*Naive Bayes* merupakan metode probabilistik pengklasifikasian sederhana berdasarkan *Teorema Bayes* dimana pengklasifikasian dilakukan melalui *training set* sejumlah data secara efisien. *Naive Bayes* mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain. *Teorema Bayes* sendiri dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes* [4].

Bentuk umum atau persamaan dari teorema *Bayes* adalah [4] :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

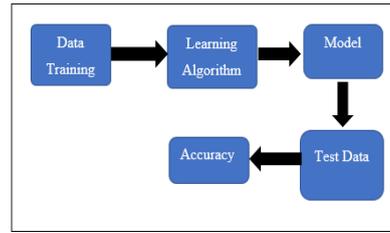
*X* : sample data yang memiliki kelas (label) yang tidak diketahui

*H* : hipotesa bahwa *X* adalah data kelas (label)

*P(H)* : peluang dari hipotesa *H*

*P(X|H)* : peluang dari data sampel *X* bila diasumsikan bahwa hipotesa benar

*P(X)* : peluang dari data sampel yang diamati  
 Pada gambar 1.1 dijelaskan tentang alur pemecahan metode klasifikasi.



**Gambar 1.1** Fase Penyelesaian Metode Klasifikasi

*Confusion Matrix* adalah salah satu cara yang digunakan untuk melakukan pengukuran terhadap nilai *accuracy* tingkat kedekatan hasil prediksi dengan hasil fakta, nilai *precision* tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, dan nilai *recall* tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Berikut tabel *confusion matrix* [3] :

**Tabel 1.1** *Confusion Matrix*

|       | Prediksi Kelas |   |                     | Total          |
|-------|----------------|---|---------------------|----------------|
|       | Negative       | Positive                                  |                     |                |
| Kelas | Negative       | True Negative (TN)<br>False Negative (FN) | False Positive (FP) | Total Negative |
|       | Positive       |   | True Positive (TP)  | Total Positive |
| Total | Total Negative | Total Positive                            |                     | Grand Total    |

Keterangan :

TP : Jumlah data positif yang terklarifikasi dengan benar oleh sistem.

TN : Jumlah data negatif yang terklarifikasi dengan benar oleh sistem

FN : Jumlah data negatif namun terklarifikasi salah oleh sistem.

FP : Jumlah data positif namun terklarifikasi salah oleh sistem.

Rumus *Accuracy* :  $Accuracy = \frac{TP}{TP+TN+FP+FN}$

Rumus *Recall* :  $Recall = \frac{TP}{TP+FN}$

Rumus *Precision* :  $Precision = \frac{TP}{(TP+FP)}$

**RapidMiner**

*RapidMiner* merupakan sebuah lingkungan *machine learning data mining, text mining, dan predictive analytics*. *RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining, text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining, termasuk operator

input,output, data *preprocessing* dan visualisasi [5].

**Penjualan**

Adapun cara-cara penjualan yang dapat dilakukan yaitu [6] :

1. Penjualan Langsung  
Merupakan cara penjualan dimana penjual langsung bertemu muka dengan calon pembeli.
2. Penjualan Tidak Langsung  
Merupakan cara penjualan dimana penjual tidak berhadapan muka secara langsung dengan calon pembeli.

**Pulsa**

Pulsa adalah media penghubung untuk berkomunikasi, sistem perhitungan dalam menentukan tarif atau biaya. Fungsi pulsa adalah sebagai satuan biaya untuk melakukan komunikasi, mengirim pesan *short message service*, dan *multimedia service* yang digunakan pada telepon selular. Pulsa juga berfungsi untuk pembayaran listrik, telepon rumah, dan PPOB (*payment poin online bank*). Pulsa juga digunakan pada alat penghubung, yaitu telepon selular [7] :

**2. METODOLOGI PENELITIAN**

**Objek Penelitian**

*Dataset* yang digunakan merupakan data penjualan pulsa bulan Juli dan Agustus 2019 pada konter RA Cell diperoleh data sebanyak 723 data kemudian untuk memudahkan proses perhitungan data penulis melakukan proses *pre-processing* sehingga diperoleh data sebanyak 600 data.

**Pengumpulan Data**

*Dataset* berjumlah 600 data yang dipakai penulis untuk pemilihan produk terlaris dalam penjualan pulsa dijadikan *training* 80% adalah sebanyak 480 *record* data, sedangkan data *testing* 20% sebanyak 180 data.

**Pengolahan Data**

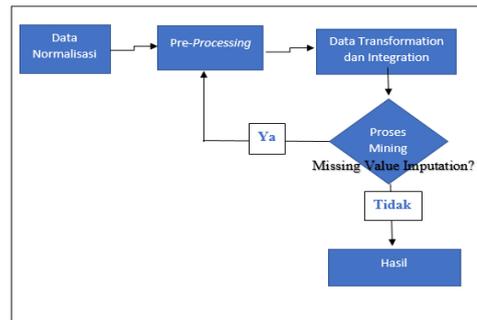
**Data Selection**

Pada tabel 1 di bawah ini merupakan pemilihan variabel sebelum dilakukan pengolahan data.

**Tabel 2.1 Data Konsumen**

| Variabel | Keterangan  | Indikator | Detail Penggunaan |
|----------|-------------|-----------|-------------------|
| X1       | No HP/Token | X         | -                 |
| X2       | Nama Produk | V         | Digunakan Atribut |
| X3       | Nominal     | V         | Digunakan Atribut |
| X4       | Harga Jual  | V         | Digunakan Atribut |

Pada gambar 1 merupakan alur *pre-processing* pengolahan data.



**Gambar 2.1 Alur Pre-Processing**

**Data Transformation dan Integration**

Pada tabel 2 di bawah ini merupakan sampel data yang diklasifikasi.

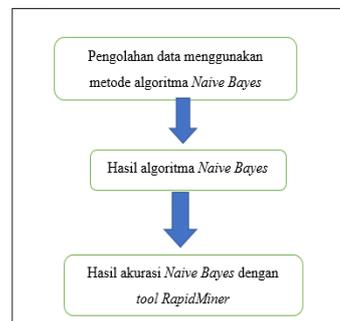
**Tabel 2.2 Sampel yang di Klasifikasi**

**Nomor dan Nama Gambar di bawah Gambar**

| No  | Nama Produk     | Nominal | Harga Jual | Peminat | Klasifikasi |
|-----|-----------------|---------|------------|---------|-------------|
| 1   | Pulsa Telkomsel | Rendah  | Murah      | Banyak  | Laris       |
| 2   | Pulsa Indosat   | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 3   | Pulsa Telkomsel | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 4   | Pulsa Indosat   | Sedang  | Sedang     | Banyak  | Laris       |
| 5   | Pulsa 3         | Tinggi  | Mahal      | Sedikit | Tidak Laris |
| 6   | Pulsa XL        | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 7   | Pulsa 3         | Sedang  | Mahal      | Banyak  | Laris       |
| 8   | Pulsa Indosat   | Tinggi  | Mahal      | Banyak  | Laris       |
| 9   | Pulsa XL        | Rendah  | Murah      | Sedikit | Tidak Laris |
| 10  | Pulsa 3         | Rendah  | Murah      | Sedikit | Tidak Laris |
| ... | ...             | ...     | ...        | ...     | ...         |
| ... | ...             | ...     | ...        | ...     | ...         |
| 598 | Pulsa Telkomsel | Rendah  | Murah      | Banyak  | Laris       |
| 599 | Pulsa Indosat   | Tinggi  | Mahal      | Banyak  | Laris       |
| 600 | Pulsa Telkomsel | Rendah  | Murah      | Banyak  | Laris       |

**Proses Mining**

Gambar 3.2 merupakan langkah pengujian data.



**Gambar 2.2 Langkah Pengujian Data**

**3. HASIL DAN PEMBAHASAN**

Hasil pengujian pada penelitian ini dilakukan dengan pengolahan data menggunakan *microsoft excel*, kemudian diuji dengan *tool* dibantu *RapidMiner* dengan data

training dan data testing yang telah dipersiapkan untuk menentukan klasifikasi Laris dan Tidak Laris pada produk penjualan pulsa. Pengujian pada penelitian ini dilakukan dengan menggunakan metode algoritma *Naive Bayes*. Dari hasil yang ada kemudian data dikategorikan dengan *variable*, *atribut* kemudian dijadikan data *training* sebanyak 80% yaitu 480 data dan data *testing* sebanyak 20% yaitu 120 data [8], dari proses tersebut kemudian dihitung untuk memperoleh *accuracy*, *precision*, dan *recall* dari algoritma *Naive Bayes* serta klasifikasi produk terlaris berdasarkan nama produk dalam penjualan pulsa.

**Perhitungan Algoritma Naive Bayes**

Untuk menentukan data yang akan diklasifikasi dengan metode *Naive Bayes* maka langkah yang harus dilakukan adalah membaca data training. Data yang akan dijadikan sebagai data *training* adalah sebanyak 480 data.

Tabel 3.1 merupakan sampel data *training*.

**Tabel 3.1 Data Training**

| No  | Nama Produk     | Nominal | Harga Jual | Peminat | Klasifikasi |
|-----|-----------------|---------|------------|---------|-------------|
| 1   | Pulsa Telkomsel | Rendah  | Murah      | Banyak  | Laris       |
| 2   | Pulsa Indosat   | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 3   | Pulsa Telkomsel | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 4   | Pulsa Indosat   | Sedang  | Sedang     | Banyak  | Laris       |
| 5   | Pulsa 3         | Tinggi  | Mahal      | Sedikit | Tidak Laris |
| ... | ...             | ...     | ...        | ...     | ...         |
| ... | ...             | ...     | ...        | ...     | ...         |
| ... | ...             | ...     | ...        | ...     | ...         |
| ... | ...             | ...     | ...        | ...     | ...         |
| 480 | Pulsa 3         | Tinggi  | Mahal      | Sedikit | Tidak Laris |

Tahap awal dalam proses algoritma *Naive Bayes* dengan tujuan untuk menentukan klasifikasi guna menentukan probabilitas dari kelas/label. Probabilitas kelas  $C_o$  "Laris" dan probabilitas  $C_i$  "Tidak Laris". Perhitungannya yaitu untuk memperoleh hasil klasifikasi Laris dan Tidak Laris dari data *training* dengan seluruh total *dataset*.

**Tabel 3.2 Probabilitas Klasifikasi**

| P(Laris/Tidak Laris) | Laris  | Tidak Laris |
|----------------------|--------|-------------|
| Laris                | 0,5604 | 0,4395      |
| Total                | 100%   |             |

Berikut merupakan tabel perhitungan *posterior* data *training*.

**Tabel 3.3 Probabilitas Nama Produk**

| P>Nama Produk)  | Klasifikasi |             | Probabilitas |             |
|-----------------|-------------|-------------|--------------|-------------|
|                 | Laris       | Tidak Laris | Laris        | Tidak Laris |
| Pulsa Telkomsel | 107         | 65          | 0,3992       | 0,3080      |
| Pulsa XL        | 38          | 39          | 0,1417       | 0,1848      |
| Pulsa Indosat   | 95          | 39          | 0,3544       | 0,1848      |
| Pulsa 3         | 29          | 68          | 0,1082       | 0,3222      |
| Total           | 269         | 211         | 100%         | 100%        |

**Tabel 3.4 Probabilitas Nominal**

| P(Nominal) | Klasifikasi |             | Probabilitas |             |
|------------|-------------|-------------|--------------|-------------|
|            | Laris       | Tidak Laris | Laris        | Tidak Laris |
| Tinggi     | 115         | 29          | 0,4275       | 0,1374      |
| Sedang     | 77          | 95          | 0,2862       | 0,4502      |
| Rendah     | 77          | 87          | 0,2862       | 0,4123      |
| Total      | 269         | 211         | 100%         | 100%        |

**Tabel 3.5 Probabilitas Harga**

| P(Harga) | Klasifikasi |             | Probabilitas |             |
|----------|-------------|-------------|--------------|-------------|
|          | Laris       | Tidak Laris | Laris        | Tidak Laris |
| Mahal    | 125         | 29          | 0,4646       | 0,1374      |
| Sedang   | 67          | 95          | 0,2490       | 0,4502      |
| Murah    | 77          | 87          | 0,2862       | 0,4123      |
| Total    | 269         | 211         | 100%         | 100%        |

**Tabel 3.6 Probabilitas Peminat**

| P(Peminat) | Klasifikasi |             | Probabilitas |             |
|------------|-------------|-------------|--------------|-------------|
|            | Laris       | Tidak Laris | Laris        | Tidak Laris |
| Banyak     | 269         | 9           | 1            | 0,0426      |
| Sedikit    | 0           | 202         | 0            | 0,9573      |
| Total      | 269         | 211         | 100%         | 100%        |

**Perhitungan Menggunakan Microsoft Excel**

Data testing pada pengujian ini berjumlah 120 data dengan. Pada tabel 3.7 dibawah ini merupakan tabel data *testing*.

**Tabel 3.7 Data Testing**

| No  | Nama Produk     | Nominal | Harga Jual | Peminat | Klasifikasi |
|-----|-----------------|---------|------------|---------|-------------|
| 1   | Pulsa Telkomsel | Rendah  | Murah      | Banyak  | Laris       |
| 2   | Pulsa Indosat   | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 3   | Pulsa Telkomsel | Sedang  | Sedang     | Sedikit | Tidak Laris |
| 4   | Pulsa Indosat   | Sedang  | Sedang     | Banyak  | Laris       |
| 5   | Pulsa 3         | Tinggi  | Mahal      | Sedikit | Tidak Laris |
| ... | ...             | ...     | ...        | ...     | ...         |
| 28  | Pulsa Telkomsel | Tinggi  | Mahal      | Banyak  | Laris       |
| ... | ...             | ...     | ...        | ...     | ...         |
| ... | ...             | ...     | ...        | ...     | ...         |
| 120 | Pulsa 3         | Tinggi  | Mahal      | Sedikit | Tidak Laris |

$$\begin{aligned}
 &= P(\text{Nama Produk} = \text{"Pulsa Telkomsel"} | \text{Laris} \\
 &\quad * \text{Nominal} = \text{"Tinggi"} | \text{Laris} * \text{Harga} = \\
 &\quad \text{"Mahal"} | \text{Laris} * \text{Peminat} = \text{"Banyak"} | \text{Laris}) * \\
 &\quad P(\text{Klasifikasi} = \text{"Laris"}) \\
 &= (0,3992 \times 0,4275 \times 0,4646 \times 1) \times (0,5604) \\
 &= 0,0444 \\
 &= P(\text{Nama Produk} = \text{"Pulsa Telkomsel"} | \text{Tidak} \\
 &\quad \text{Laris} * \text{Nominal} = \text{"Tinggi"} | \text{Tidak Laris} *
 \end{aligned}$$

$$\begin{aligned} \text{Harga} &= \text{"Mahal"} | \text{Tidak Laris} * \text{Peminat} = \\ &= \text{"Banyak"} | \text{Tidak Laris} * P(\text{Klasifikasi} = \\ &= \text{Banyak} | \text{"Tidak Laris"}) \\ &= (0,3080 \times 0,1374 \times 0,1374 \times 0,0426) \times (0,4395) \\ &= 0,0001 \end{aligned}$$

Membandingkan variabel Laris dan Tidak Laris

Berdasarkan perhitungan akhir dengan mengalikan nilai peluang dari kasus yang diangkat, kita melihat perbandingan hasil variabel Laris dan Tidak Laris dibawah ini :

Diketahui :

$C_o$  = Laris

$C_i$  = Tidak Laris

Ditanyakan : Hasil perbandingan  $P(X|C_o) * P(C_o)$  dan  $P(X|C_i) * P(C_i)$  ?

Jawab :

$$P(X|C_o) * P(C_o) = P(X|C_i) * P(C_i)$$

$$0,0444 = 0,0001$$

$$0,0444 > 0,0001$$

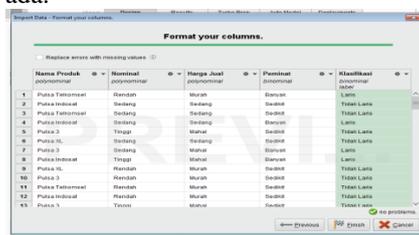
Jadi data uji  $P(X|C_o) * P(C_o) > P(X|C_i) * P(C_i)$  nilai Laris lebih tinggi nilainya dibandingkan dengan nilai tidak Laris. Sehingga dapat disimpulkan bahwa produk Pulsa Telkomsel masuk dalam klasifikasi "Laris".

### Perhitungan Metode Naive Bayes Menggunakan RapidMiner

Dalam rangkaian proses data mining menggunakan tool RapidMiner data yang digunakan adalah data yang telah di transformasi kedalam format Microsoft Excel. Pengujian ini untuk mengetahui nilai akurasi dan prediksi data. Pengujian dilakukan dengan algoritma Naive Bayes dan data yang digunakan sebanyak 480 data training dan 120 data testing.

### Import Data Training

Data training ini digunakan untuk latihan membentuk sebuah model classifier untuk prediksi kelas data baru yang belum pernah ada.



Gambar 3.1 Import Data Training

### Import Data Testing

Data testing ini digunakan untuk mengukur sejauh mana prediction berhasil melakukan prediksi kualitas produk dengan benar.



Gambar 3.2 Import Data Testing

### Proses Pengujian

Proses klasifikasi pada RapidMiner menggunakan metode Naive Bayes untuk membandingkan data training dan data testing yang sudah diketahui.



Gambar 3.3 Model Utama Naive Bayes pada RapidMiner

### Hasil Apply Model

Hasil dari apply model dapat dilihat pada tabel nomor 2 data testing yang sudah mendapatkan hasil prediksi "Layak" dengan menggunakan RapidMiner.

Berikut tabel hasil pengujian data training dengan data testing menggunakan RapidMiner dapat dilihat pada tabel di bawah ini :

| No | Klasifikasi | prediksi    | confusion | confusion | Nama P.         | Nonimal | Harga J. | Peminat | L | T |
|----|-------------|-------------|-----------|-----------|-----------------|---------|----------|---------|---|---|
| 1  | Laris       | Laris       | 0,999     | 0,999     | Pulsa Telkomsel | Murah   | 1000     | 1000    | 1 | 1 |
| 2  | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa Indosat   | Sedang  | 1000     | 1000    | 1 | 1 |
| 3  | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa Indosat   | Sedang  | 1000     | 1000    | 1 | 1 |
| 4  | Laris       | Laris       | 0,999     | 0,999     | Pulsa 3         | Mahal   | 1000     | 1000    | 1 | 1 |
| 5  | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa 2         | Sedang  | 1000     | 1000    | 1 | 1 |
| 6  | Laris       | Laris       | 0,999     | 0,999     | Pulsa XL        | Sedang  | 1000     | 1000    | 1 | 1 |
| 7  | Laris       | Laris       | 0,999     | 0,999     | Pulsa XL        | Sedang  | 1000     | 1000    | 1 | 1 |
| 8  | Laris       | Laris       | 0,999     | 0,999     | Pulsa XL        | Sedang  | 1000     | 1000    | 1 | 1 |
| 9  | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa 3         | Mahal   | 1000     | 1000    | 1 | 1 |
| 10 | Laris       | Laris       | 0,999     | 0,999     | Pulsa 2         | Sedang  | 1000     | 1000    | 1 | 1 |
| 11 | Laris       | Laris       | 0,999     | 0,999     | Pulsa 2         | Sedang  | 1000     | 1000    | 1 | 1 |
| 12 | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa XL        | Sedang  | 1000     | 1000    | 1 | 1 |
| 13 | Tidak Laris | Tidak Laris | 0,999     | 0,999     | Pulsa 2         | Sedang  | 1000     | 1000    | 1 | 1 |
| 14 | Laris       | Laris       | 0,999     | 0,999     | Pulsa Telkomsel | Murah   | 1000     | 1000    | 1 | 1 |
| 15 | Laris       | Laris       | 0,999     | 0,999     | Pulsa Telkomsel | Murah   | 1000     | 1000    | 1 | 1 |

Gambar 3.4 Hasil Prediksi pada RapidMiner Pada gambar 3.5 merupakan PerformanceVector menggunakan metode algoritma Naive Bayes pada RapidMiner.

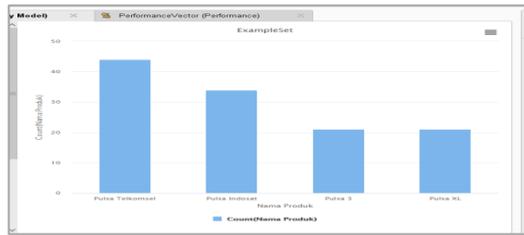
```

PerformanceVector:
accuracy: 97,90%
ConfusionMatrix:
True: Laris Tidak Laris
Laris: 74 3
Tidak Laris: 0 43
precision: 100,00% (positive class: Tidak Laris)
ConfusionMatrix:
True: Laris Tidak Laris
Laris: 74 3
Tidak Laris: 0 43
recall: 93,48% (positive class: Tidak Laris)
ConfusionMatrix:
True: Laris Tidak Laris
Laris: 74 3
Tidak Laris: 0 43
AUC (optimistic): 0,959 (positive class: Tidak Laris)
AUC: 0,955 (positive class: Tidak Laris)
AUC (pessimistic): 0,952 (positive class: Tidak Laris)
    
```

Gambar 3.5 Performanvector Naive Bayes pada RapidMiner

Pengujian dengan metode algoritma Naive Bayes menggunakan RapidMiner menghasilkan produk penjualan pulsa terlaris berdasarkan nama produk yang ada pada

konter RA Cell adalah produk Pulsa Telkomsel. Gambar 3.6 merupakan grafik produk terlaris berdasarkan nama produk



**Gambar 3.6** Grafik Produk Terlaris Berdasarkan Nama Produk

Pada gambar 3.7 dibawah ini dijelaskan tentang tabel jumlah hasil penjualan pulsa menurut nama produk yang ada pada konter RA Cell. Hasil Produk terlaris yaitu Pulsa Telkomsel dengan jumlah 44 data.

| Index | Nominal value   | Absolute count | Fraction |
|-------|-----------------|----------------|----------|
| 1     | Pulsa Telkomsel | 44             | 0.367    |
| 2     | Pulsa Indosat   | 34             | 0.283    |
| 3     | Pulsa 3         | 21             | 0.175    |
| 4     | Pulsa XL        | 21             | 0.175    |

**Gambar 3.7** Hasil Data Produk Laris Berdasarkan Nama Produk

Untuk mendapatkan nilai *accuracy* tingkat kedekatan hasil prediksi dengan hasil fakta, *precision* tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, dan *recall* tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Pada gambar 4.8 menjelaskan tentang hasil perhitungan *accuracy* dari algoritma *Naive Bayes* dilakukan dengan cara menjumlahkan TP + TN dibagi jumlah total data *testing* yang diuji.

accuracy: 97,50%

|                   | true Laris | true Tidak Laris | class precision |
|-------------------|------------|------------------|-----------------|
| pred. Laris       | 74         | 3                | 96.10%          |
| pred. Tidak Laris | 0          | 43               | 100.00%         |
| class recall      | 100.00%    | 93.48%           |                 |

**Gambar 3.8** Hasil *Accuracy* Algoritma *Naive Bayes* pada *RapidMiner*

$$Accuracy = \frac{TP}{TP + TN + FP + FN} * 100\%$$

$$= \frac{74 + 43}{74 + 43 + 0 + 3} * 100\%$$

$$= \frac{117}{120} * 100\%$$

$$= 97,50\%$$

Pada gambar 3.9 menjelaskan tentang hasil perhitungan *precision* dari algoritma *Naive Bayes* bahwa tingkat *Precision* mencapai 100,00%

precision: 100.00% (positive class: Tidak Laris)

|                   | true Laris | true Tidak Laris | class precision |
|-------------------|------------|------------------|-----------------|
| pred. Laris       | 74         | 3                | 96.10%          |
| pred. Tidak Laris | 0          | 43               | 100.00%         |
| class recall      | 100.00%    | 93.48%           |                 |

**Gambar 3.9** Hasil *Precision* Algoritma *Naive Bayes* pada *RapidMiner*

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$= \frac{74}{74 + 0} * 100\%$$

$$= \frac{74}{74} * 100\%$$

$$= 100\%$$

Pada gambar 4.10 menjelaskan tentang hasil perhitungan *Recall* dari algoritma *Naive Bayes* bahwa tingkat *Recall* mencapai 93,48%

recall: 93.48% (positive class: Tidak Laris)

|                   | true Laris | true Tidak Laris | class precision |
|-------------------|------------|------------------|-----------------|
| pred. Laris       | 74         | 3                | 96.10%          |
| pred. Tidak Laris | 0          | 43               | 100.00%         |
| class recall      | 100.00%    | 93.48%           |                 |

**Gambar 3.10** Hasil *Recall* algoritma *Naive Bayes* Pada *RapidMiner*

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$= \frac{43}{43 + 3} * 100\%$$

$$= \frac{43}{46} * 100\%$$

$$= 93,48\%$$

#### 4. SIMPULAN

Hasil pembahasan yang telah diuraikan oleh penulis tentang penerapan Algoritma *Naive Bayes* untuk menentukan klasifikasi produk terlaris pada penjualan pulsa, maka penulis dapat mengambil kesimpulan yaitu sebagai berikut :

1. Berdasarkan hasil perhitungan yang telah dilakukan oleh penulis dengan perhitungan menggunakan Algoritma *Naive Bayes* yang dihitung secara manual, *Microsoft Excel*, dan tool *RapidMiner* menghasilkan data yang sama sehingga data *mining* algoritma metode *Naive Bayes* sangat tepat digunakan untuk menghitung

- klasifikasi data penjualan pulsa karena menghasilkan data yang akurat.
2. Berdasarkan hasil perhitungan menggunakan metode *Naive Bayes* produk terlaris dalam penjualan pulsa adalah Pulsa Telkomsel.
  3. Berdasarkan hasil perhitungan Metode *Naive Bayes* menggunakan *RapidMiner* nilai *accuracy* : 97,50%, *Precision* : 100,00%, dan *Recall* : 93,48%.

## 5. DAFTAR PUSTAKA

- [1] [N. Dicky, E. Kamil dan R. Mukhlis, "Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Untuk Mengetahui Minat Beli Pelanggan Terhadap Kartu Internet XL," Jurnal Ilmiah Saintikom, vol. 15, p. 2, Mei 2016.
- [2] L. W. Supriyadi, "Identifikasi Pemakaian Energi Listrik Pelanggan Menggunakan Metode Boosting Naive Bayes," Makassar, 2017.
- [3] Y. Mardi, "Klasifikasi Menggunakan Algoritma C4.5," Jurnal Edik Informatika, vol. 2, p. 2, 2019.
- [4] S. Alfa, "Implementasi Metode Klasifikasi Naive Bayes Dalam Mengkalsifikasi Besarnya penggunaan Listrik Rumah Tangga," Citec Journal, vol. 2 No. 3, Mei-Juni 2015.
- [5] N. Odi dan Noval Salim, "Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran," Jurnal Teknologi Informasi dan Komunikasi, 2018.
- [6] E. P. Yulia, W. A. Winda dan M. Nurkhamid, "Perancangan Sistem Penjualan Pulsa," Jurnal Ilmiah Fifo, vol. VIII, p. 2, 2016.
- [7] D. Kurniawan, A. Hijriani dan F. T. Hartawan, "Rancang Bangun Aplikasi Agen Pulsa Elektronik Berbasis Android," Jurnal Komputisi, vol. 3, p. 2, 2015.
- [8] F. C. Apriliya, S. Ristu dan W. S. Sari, "Penentuan Terbaik Pada Metode Naive Bayes Classifier Dalam Menentukan Status Gizi Balita Dengan Mempertimbangkan Independensi Parameter," Jurnal ITSMART, p. 4 No 1, 2015.