

PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENENTUKAN KLASIFIKASI PRODUK TERLARIS PADA PENJUALAN PULSA

Ismasari Nawangsih¹⁾, Asti Setyaningsih²⁾

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
ismasari.n@pelitabangsa.ac.id

Disetujui, 29 Juni 2020

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 peneliti mengklasifikasi menjadi Laris dan Tidak Laris. Tujuannya dapat mengetahui penerapan data mining dengan menggunakan algoritma Naive 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 RapidMiner dan algoritma Naive Bayes merupakan metode klasifikasi yang banyak digunakan karena sederhana dan akurasi yang tinggi dalam mengklasifikasi data. Berdasarkan hasil penelitian yang telah dilakukan, jenis produk terlaris pada penjualan pulsa menurut nama produk adalah Pulsa Telkomsel. Tingkat akurasi klasifikasi dengan metode Naive Bayes menghasilkan nilai Accuracy sebesar 97,50%, nilai Precision 100%, dan nilai Recall 93,48% sehingga metode Naive Bayes merupakan metode yang cukup baik dalam penelitian ini.

Kata Kunci: Pulsa, Klasifikasi, Data Mining, Naive Bayes

Abstract

This research is motivated by the progress of the development of communication technology and information is very fast and increasingly cheap so that makes the community for mobile phone pulses become a mandatory requirement at the present time. Of various pulse products available at the counter RA Cell Pulses Telkomsel, Pulses XL, Pulses Indosat, and Pulses 3 the authors classify as bestseller and non-sellers. The goal is to find out the implementation of data mining using the Naive 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 RapidMiner application and the Naive Bayes algorithm is a classification method that is widely used because of its simple and high accuracy in classifying data. Based on the result of research that has been done, the type of product that is most restricted to the sale of pulses by product name is Telkomsel Pulses. The level of classification accuracy with the naive Bayes method produces an accuracy value of 97,50%, a precision value of 100%, and a recall value of 93,48% so the Naive Bayes method is good method in this study.

Keywords : Pulses, Classification, Data Mining, Naive 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 pulsaelektrik 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 larispada penjualan pulsa yaitu dapat menghasilkan data yang memiliki tingkat akurasi yang dapat dimanfaatkan untuk memprediksi kelas dari kasus yang baru. Dengan menggunakan *tool RapidMiner* dalam proses perhitungannya sehingga dapatdiketahui 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.

2. Tinjauan Studi

2.1 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].

2.2 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].

2.3 Naive Bayes

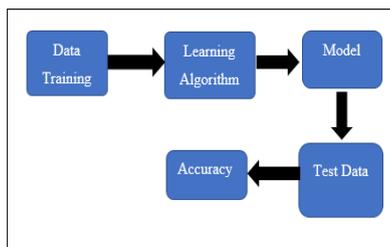
Naive Bayes merupakan metode probabilistik pengklasifikasian sederhana berdasarkan *Teorema Bayes* dimana pengklasifikasian dilakukan melalui *training set* sejumlah data secara efisien. *Naive Bayes* mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan 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] :

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

Keterangan

- 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* biladiasumsikan bahwa hipotesa benar
- P(X)* : peluang dari data sampel yang diamati

Pada gambar 2.1 dijelaskan tentang alur pemecahan metode klasifikasi.



Gambar 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. *Confusion Matrix*

		Prediksi Kelas		
		Negative	Positive	Total
Kelas	Negative	True Negative (TN)	False Positive (FP)	Total Negative
	Positive	False Negative (FN)	True Positive (TP)	Total Positive
Total		Total Predicted Negative	Total Predicted 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

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

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

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

2.4 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].

2.5 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.

2.6 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]

3. Desain Penelitian/ Metodologi

3.1 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 peneliti melakukan proses pre-processing sehingga diperoleh data sebanyak 600 data.

3.2 Pengumpulan Data

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

3.3 Pengolahan Data

3.3.1 Data Selection

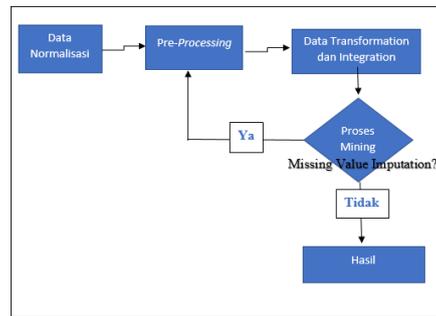
Pada tabel 1 di bawah ini merupakan pemilihan variabel sebelum dilakukan pengolahan data.

Tabel 2. 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

3.3.2 Pre-Processing

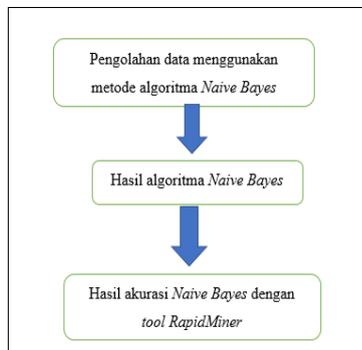
Pada gambar 3.1 merupakan alur pre- processing pengolahan data.



Gambar 2. Alur Pre-Processing

3.3.3 Data Transformation dan Integrasi

Pada tabel 2 di bawah ini merupakan sampel data yang diklasifikasi



Gambar 3. Proses Mining

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

Gambar 4. Langkah Pengujian Data

Bentuk umum atau persamaan dari teorema Bayes adalah :

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

Keterangan :

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 biladiasumsikan bahwa hipotesa benar

$P(X)$: peluang dari data sampel yang diamati
Perhitungan kriteria, diperlukan persamaan :

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

Menghitung $P(C_i)$ yang merupakan probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan :

$$P(C_i) = \frac{S_i}{S}$$

Dimana:

S_i : Jumlah data training dari kategori

S : Jumlah total data training.

4. 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.

4.1. 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 4.1 merupakan sampel data *training*.

Tabel 3. 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_0 “Laris” dan probabilitas C_i “Tidak Laris”. Perhitungannya yaitu untuk

memperoleh hasil klasifikasi Laris dan Tidak Laris dari data *training* dengan seluruh total *dataset*.

Diketahui :

$$\text{Kelas } C_0 \text{ "Laris"} = 269$$

$$\text{Kelas } C_i \text{ "Tidak Laris"} = 211$$

Ditanya Probabilitas kelas C_0 "Laris" dan Probabilitas C_i "Tidak Laris" ?Jawab :

Probabilitas C_0 "Laris" adalah :

$$P(C) = \frac{269}{480} = 0,5604$$

Probabilitas C_i "Tidak Laris" adalah :

$$P(C) = \frac{211}{480} = 0,4395$$

Pada tabel 4.2 merupakan tabel probabilitas klasifikasi.

Tabel 4. Probabilitas Klasifikasi

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

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

Tabel 5. 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%

a. Menghitung nilai P (Nama Produk)

$$= X|(C_0) P (Nama Produk) = X|(C_0)$$

P (Nama Produk = "Pulsa

$$\text{Telkomsel"} | \text{Klasifikasi} = \text{"Laris"} P (Nama Produk = 107/269) = 0,3992$$

Menghitung nilai P (Nama Produk)

$$= X|(C_i)$$

P (Nama Produk = "Pulsa Telkomsel" | Klasifikasi = "Tidak Laris")

$$P (Nama Produk = 65/211) = 0,3080$$

b. Menghitung nilai P (Nama Produk)

$$= X|(C_0)$$

P (Nama Produk = "Pulsa XL" | Klasifikasi = "Laris")

$$P (Nama Produk = 38/269) = 0,1417$$

Menghitung nilai P (Nama Produk)

$$= X|(C_i)$$

P (Nama Produk = "Pulsa XL" |

Klasifikasi "Tidak Laris")

$$P (Nama Produk = 39/211) = 0,1848$$

c. Menghitung nilai P (Nama Produk)

- = $X|(Co)$
- P (Nama Produk = “Pulsa Indosat” | Klasifikasi = “Laris”)
- P (Nama Produk = 95/269) = 0,3544 Menghitung nilai P (Nama Produk)
- = $X|(Ci)$
- P (Nama Produk = “Pulsa Indosat” | Klasifikasi “Tidak Laris”)
- P (Nama Produk = 39/211) = 0,1848
- d. Menghitung nilai P (Nama Produk)
- = $X|(Co)$
- P (Nama Produk = “Pulsa 3” |
- Klasifikasi = “Laris”)
- P (Nama Produk = 29/269) = 0,1082 Menghitung nilai P (Nama Produk)
- = $X|(Ci)$
- P (Nama Produk = “Pulsa 3” | Klasifikasi “Tidak Laris”)
- P (Nama Produk = 68/211) = 0,3222

Tabel 6. 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%

- a. Menghitung nilai P (Nominal) = $X|(Co)$ P (Nominal = “Tinggi” | Klasifikasi = “Laris”) P (Nominal = 115/269) = 0,4275 Menghitung nilai P (Nominal) = $X|(Ci)$ P (Nominal = “Tinggi” | Klasifikasi = “Tidak Laris”) P (Nominal = 29/211) = 0,1374
- b. Menghitung nilai P (Nominal) = $X|(Co)$ P (Nominal = “Sedang” | Klasifikasi = “Laris”) P (Nominal = 77/269) = 0,2862 Menghitung nilai P (Nominal) = $X|(Ci)$ P (Nominal = “Sedang” | Klasifikasi “Tidak Laris”) P (Nominal = 95/211) = 0,4502
- c. Menghitung nilai P (Nominal) = $X|(Co)$ P (Nominal = “Rendah” | Klasifikasi = “Laris”) P (Nominal = 77/269) = 0,2862 Menghitung nilai P (Nominal) = $X|(Ci)$ P (Nominal = “Rendah” | Klasifikasi = “Tidak Laris”) P (Nominal = 87/211) = 0,4123

Tabel 7. 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%

- a. Menghitung nilai P (Harga) = $X|(Co)$ P (Harga = “Mahal” | Klasifikasi = “Laris”) P (Harga = 125/269) = 0,4646 Menghitung nilai P (Nama Harga) = $X|(Ci)$ P (Harga = “Mahal” | Klasifikasi “Tidak Laris”) P (Harga = 29/211) = 0,1374
- b. Menghitung nilai P (Harga) = $X|(Co)$ P (Harga = “Sedang” | Klasifikasi = “Laris”) P (Harga = 67/269) =

0,2490 Menghitung nilai $P(\text{Harga} = X|C_i) P(\text{Harga} = \text{"Sedang"} | \text{Klasifikasi} = \text{"Tidak Laris"}) P(\text{Harga} = 95/211) = 0,4502$

- c. Menghitung nilai $P(\text{Harga} = X|C_0) P(\text{Harga} = \text{"Murah"} | \text{Klasifikasi} = \text{"Laris"}) P(\text{Harga} = 77/269) = 0,2862$ Menghitung nilai $P(\text{Harga} = X|C_i) P(\text{Harga} = \text{"Murah"} | \text{Klasifikasi} = \text{"Tidak Laris"}) P(\text{Harga} = 87/211) = 0,4123$

Tabel 8. 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%

- a. Menghitung nilai $P(\text{Peminat} = X|C_0) P(\text{Peminat} = \text{"Banyak"} | \text{Klasifikasi} = \text{"Laris"}) P(\text{Peminat} = 269/269) = 1$ Menghitung nilai $P(\text{Peminat} = X|C_i) P(\text{Peminat} = \text{"Banyak"} | \text{Klasifikasi} = \text{"Tidak Laris"}) P(\text{Peminat} = 9/211) = 0,0426$
- b. Menghitung nilai $P(\text{Peminat} = X|C_0) P(\text{Peminat} = \text{"Sedikit"} | \text{Klasifikasi} = \text{"Laris"}) P(\text{Peminat} = 0/269) = 0$ Menghitung nilai $P(\text{Peminat} = X|C_i) P(\text{Peminat} = \text{"Sedikit"} | \text{Klasifikasi} = \text{"Tidak Laris"}) P(\text{Peminat} = 202/211) = 0,9573$

4.2. Perhitungan Menggunakan Microsoft Excel

Data testing pada pengujian ini berjumlah 120 data dengan. Pada tabel 4.7 dibawah ini merupakan tabel data testing.

Tabel 9. 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

$$= P(\text{Nama Produk} = \text{"Pulsa Telkomsel"} | \text{Laris} * \text{Nominal} = \text{"Tinggi"} | \text{Laris} * \text{Harga} = \text{"Mahal"} | \text{Laris} * \text{Peminat} = \text{"Banyak"} | \text{Laris}) * P(\text{Klasifikasi} = \text{"Laris"} | \text{Banyak} | \text{"Laris"})$$

$$= (0,3992 * 0,4275 * 0,4646 * 1) * (0,5604) = 0,0444$$

$$= P(\text{Nama Produk} = \text{"Pulsa Telkomsel"} | \text{Tidak Laris} * \text{Nominal} = \text{"Tinggi"} | \text{Tidak Laris} * \text{Harga} = \text{"Mahal"} | \text{Tidak Laris} * \text{Peminat} = \text{"Banyak"} | \text{Tidak Laris}) * P(\text{Klasifikasi} = \text{"Banyak"} | \text{"Tidak Laris"})$$

$$= (0,3080 * 0,1374 * 0,1374 * 0,0426) * (0,4395) = 0,0001$$

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

$$C_0 = \text{Laris}$$

$$C_i = \text{Tidak Laris}$$

Ditanyakan: Hasil perbandingan $P(X|(C_0) * P(C_0))$ dan $P(X|(C_i) * P(C_i))$?

Jawab :

$$P(X|(C_0) * P(C_0)) = P(X|(C_i) * P(C_i)) \quad 0,0444 = 0,0001 \text{ ----- } 0,0444 > 0,0001$$

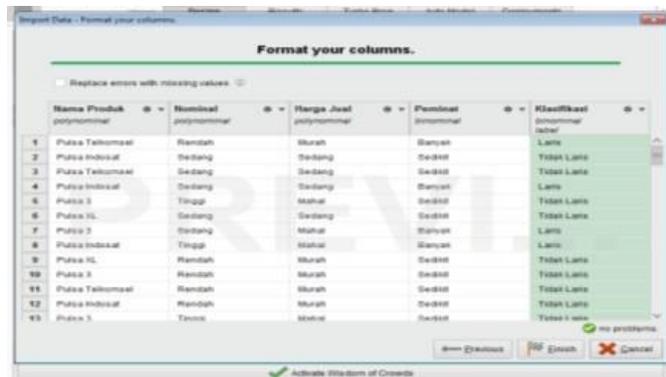
Jadi data uji $P(X|(C_0) * P(C_0)) > P(X|(C_i) * P(C_i))$ nilai Laris lebih tinggi nilainya dibandingkan dengan nilai tidak Laris. Sehingga dapat disimpulkan bahwa produk Pulsa Telkomsel masuk dalam klasifikasi “Laris”.

4.3. 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.

a. Import Data Training

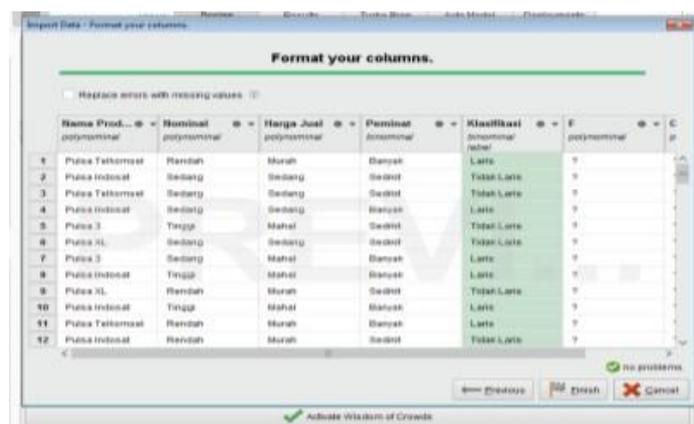
Data training ini digunakan untuk latihan membentuk sebuah model classifier untuk prediksi kelas data baru yang belum pernah ada.



Gambar 5. Import Data Training

b. Import Data Testing

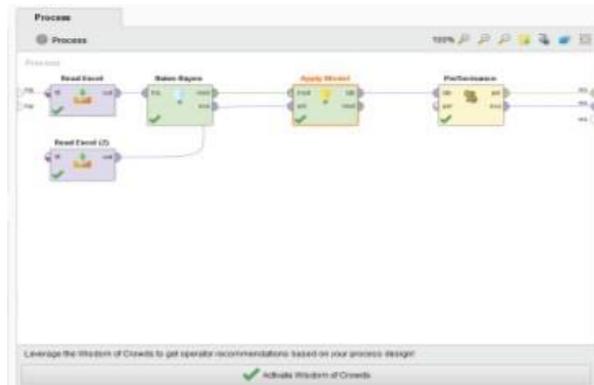
Data testing ini digunakan untuk mengukur sejauh mana prediction berhasil melakukan prediksi kualitas produk dengan benar



Gambar 6. Import Data Testing

c. Proses Pengujian

Proses klasifikasi pada RapidMiner menggunakan metode Naive Bayes untuk membandingkan data training dan data testing yang sudah diketahui. Gambar 4.3 merupakan gambar Model Utama Naive Bayes dengan RapidMiner.



Gambar 7. Model Utama *Naive Bayes* pada *RapidMiner*

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

Gambar 8. Hasil Prediksi pada *RapidMiner*

Pada gambar 4.5 merupakan *PerformanceVector* menggunakan metode algoritma *Naive Bayes* pada *RapidMiner*.

PerformanceVector

```

PerformanceVector:
accuracy: 97.50%
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)
    
```

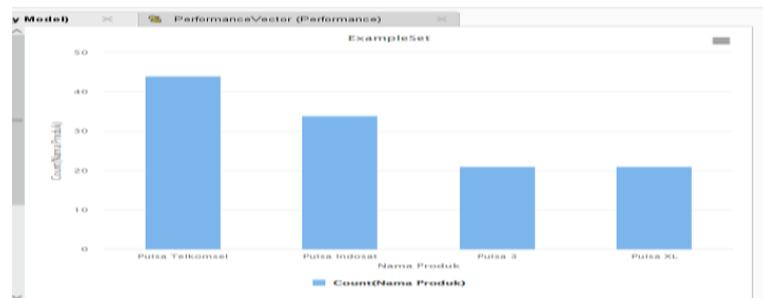
Gambar 9. *PerformanceVector* *Naive Bayes* pada *RapidMiner*

Pengujian dengan metode algoritma *Naive Bayes* menggunakan *Rapid Miner* menghasilkan produk penjualan pulsa terlaris berdasarkan nama produk yang ada pada konter RA Cell adalah produk Pulsa Telkomsel. Gambar 4.6 merupakan grafik produk terlaris berdasarkan nama produk

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 10. Grafik Produk Terlaris Berdasarkan Nama Produk

Pada gambar 4.7 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.



Gambar 11. 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 12. Hasil Accuracy Algoritma Naive Bayes pada Rapid Miner

Pada gambar 4.9 menjelaskan tentang hasil perhitungan *precision* dari algoritma *Naive Bayes* bahwa tingkat *Precision* mencapai 100,00%

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} * 100\% \\
 &= \frac{74}{74 + 0} * 100\% \\
 &= \frac{74}{74} * 100\% \\
 &= 100\%
 \end{aligned}
 \qquad
 \begin{aligned}
 Accuracy &= \frac{TP}{TP + TN + FP + FN} * 100\% \\
 &= \frac{74 + 43}{74 + 43 + 0 + 3} * 100\% \\
 &= \frac{117}{120} * 100\% \\
 &= 97,50\%
 \end{aligned}$$

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 13. Hasil *Precision* Algoritma *Naive Bayes* pada *RapidMiner*

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%	

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{43}{43 + 3} * 100\% \\
 &= \frac{43}{46} * 100\% \\
 &= 93,48\%
 \end{aligned}$$

Gambar 14. Hasil *Recall* algoritma *Naive Bayes* Pada *RapidMiner*

5. Kesimpulan

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

1. Berdasarkan hasil perhitungan yang telah dilakukan oleh peneliti 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%.

Daftar Pustaka

- 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.
- L. W. Supriyadi, "Identifikasi Pemakaian Energi Listrik Pelanggan Menggunakan Metode Boosting Naive Bayes," Makassar, 2017.
- Y. Mardi, "Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, vol. 2, p. 2, 2019.
- S. Alfa, "Implementasi Metode Klasifikasi Naive Bayes Dalam Mengkalsifikasi Besarnya penggunaan Listrik Rumah Tangga," *CitecJournal*, vol. 2 No. 3, Mei-Juni 2015.
- 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.
- E. P. Yulia, W. A. Winda dan M. Nurkhamid, "Perancangan Sistem Penjualan Pulsa," *Jurnal Ilmiah Fifo*, vol. VIII, p. 2, 2016.
- D. Kurniawan, A. Hijriani dan F. T. Hartawan, "Rancang Bangun Aplikasi Agen Pulsa Elektronik Berbasis Android," *Jurnal Komputisi*, vol. 3, p. 2, 2015.
- 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.