

Klasifikasi Penjualan *Provider* Pulsa di Kecamatan Masbagik Lombok Timur Menggunakan Metode *Naïve Bayes*

Nurul Musfita^{1*}, Nurul Fitriyani², Zulhan Widya Baskara³

^{1,2,3}Jurusan Matematika, Fakultas MIPA, Universitas Mataram, Mataram, 83115, Indonesia

*Corresponding author, email: nurul.musfita18@gmail.com

Abstract

The rapid development of technology causes the use of mobile phones and the need for pulses to increase. East Lombok is the area with the largest population in NTB and high users of information technology. East Lombok also has an internet network or smooth communication signal, which shows that there are many providers in the area. To see the types of providers that are widely used in Masbagik District, East Lombok, taking into account the largest population, a classification is made of whether these providers are in demand or not using the *Naïve Bayes* method. This study aims to determine the classification results and the accuracy of the sales classification of credit providers. The data is split into two categories: training data (90%) and testing data (10%). According to the findings of the study, 225 of the 309 testing data were correctly classified. The resulting APER value is 27.2%, which indicates that the accuracy of the classification results using the *Naïve Bayes* method is 72.8%. An AUC value of 0.804 was also obtained, which means that the accuracy of the classification of selling pulse providers that are in demand, moderately in demand, and not in demand was sufficient.

Keywords: Accuracy, Classification, *Naïve Bayes*, Provider.

Abstrak

Perkembangan teknologi yang pesat menyebabkan penggunaan *handphone* dan kebutuhan pulsa meningkat. Lombok Timur menjadi daerah dengan penduduk terbanyak di NTB dan pengguna teknologi informasi yang tinggi. Lombok Timur juga memiliki jaringan internet atau sinyal komunikasi lancar, yang menunjukkan ada banyak *provider* di wilayah tersebut. Untuk melihat jenis *provider* yang banyak digunakan di Kecamatan Masbagik Lombok Timur dengan mempertimbangkan jumlah penduduk terbanyak, dilakukan pengklasifikasian apakah *provider* tersebut laris atau tidak menggunakan metode *Naïve Bayes*. Penelitian ini bertujuan menentukan hasil klasifikasi dan ketepatan klasifikasi penjualan *provider* pulsa. Data dibagi menjadi dua jenis: data *training* (90%) dan data *testing* (10%). Berdasarkan hasil penelitian yang dilakukan, terdapat 225 data berhasil diklasifikasikan dengan benar dari 309 data *testing* yang digunakan. Dihasilkan nilai APER sebesar 27.2%, yang menunjukkan bahwa ketepatan hasil klasifikasi menggunakan metode *Naïve Bayes* adalah 72.8%. Didapatkan juga nilai AUC sebesar 0.804 yang berarti ketepatan klasifikasi penjualan *provider* pulsa laris, cukup laris, dan tidak laris sudah cukup.

Kata Kunci: Akurasi, Klasifikasi, *Naïve Bayes*, Provider.

1. Pendahuluan

Manusia merupakan makhluk sosial yang memiliki kebutuhan dalam bersosialisasi, dimana proses sosialisasi membutuhkan proses komunikasi. Perkembangan teknologi yang pesat menyebabkan proses komunikasi dan informasi dapat dilakukan dengan menggunakan *handphone*, laptop, komputer, dan lainnya. Meningkatnya kemudahan dalam mengakses teknologi digital telah mengubah perilaku konsumen di Indonesia, dimana penggunaan *smartphone* di Indonesia meningkat secara fantastis. Hal ini menyebabkan terjadinya peningkatan kompleksitas jalur informasi dan penggalian informasi yang dapat mempengaruhi aktivitas konsumen dalam penggunaan teknologi digital [1].

Dalam suatu komunikasi dan informasi, jaringan internet atau sinyal menjadi kebutuhan utama. Menurut data Dukcapil Provinsi NTB tahun 2021, Lombok Timur menjadi kabupaten dengan jumlah penduduk terbanyak yaitu 1.334.499 penduduk. Dengan banyaknya penduduk, pertumbuhan pengguna internet juga meningkat. Pada tahun 2020 juga, tidak terdapat daerah di Kabupaten Lombok Timur yang kurang sinyal komunikasi. Berdasarkan hal tersebut, diketahui bahwa ada banyak *provider* atau penyedia layanan komunikasi yang ada di Lombok Timur. Hal ini selaras dengan data BPS Lombok Timur 2021, yaitu persentase pengguna teknologi informasi khususnya telepon seluler terbilang tinggi sebesar 87.15%.

Suatu perangkat elektronik dapat berkomunikasi satu sama lain melalui internet, dimana hal tersebut dapat diakses menggunakan pulsa dan kuota. Hal ini menyebabkan kebutuhan dan penggunaan pulsa elektrik semakin meningkat. Salah satu pertimbangan pengguna jasa telekomunikasi adalah kecepatan internet dari suatu *provider* seluler [2]. Untuk melihat jenis *provider* yang paling banyak digunakan, akan dilakukan pengklasifikasian untuk memprediksi apakah *provider* tersebut laris atau tidak laris di suatu wilayah tertentu dengan menggunakan data pembelian pulsa untuk melihat efisiensi kegiatan bauran promosi di suatu daerah.

Naïve Bayes Classifier merupakan sebuah pengklasifikasian dengan menggunakan metode probabilitas dan statistik. *Naïve Bayes* memiliki beberapa kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana, serta berakurasi tinggi [3]. Digunakan *software R* untuk pengklasifikasiannya karena bersifat *open source*, termasuk untuk melakukan operasi *machine learning* seperti klasifikasi dan regresi.

Penelitian ini didasarkan pada penelitian sebelumnya oleh Nawangsih dan Setyaningsih pada tahun 2020 tentang Penerapan Algoritma *Naïve Bayes* Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. Hasil klasifikasi dan akurasi data terhadap stok penjualan pulsa menghasilkan nilai *Accuracy* 97,50% [4]. Adapun pada penelitian Devita *et al.* pada tahun 2018 tentang Perbandingan Kinerja Metode *Naive Bayes* dan *K-Nearest Neighbor* Untuk Klasifikasi Artikel Berbahasa Indonesia. Hasil yang diperoleh dalam mengklasifikasikan dokumen/artikel secara

otomatis dan akurat adalah *Naïve Bayes* memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 40% [5].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menentukan hasil klasifikasi dan ketepatan hasil klasifikasi penjualan *provider* pulsa menggunakan metode *Naïve Bayes*. Hal ini dilakukan untuk membantu operator seluler atau pemilik konter dalam mengklasifikasikan produk guna mengetahui produk yang paling laris agar operator dapat lebih maksimal dalam memasarkan produk dan dapat menentukan stok produk yang paling banyak diminati oleh masyarakat agar produk selalu tersedia ketika ada pembeli.

2. Material dan Metode

2.1 Data Mining

Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. *Data mining* bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan [6].

2.2 Training Set dan Testing Set

Training set adalah sekumpulan data yang digunakan dalam mentraining data. *Training set* digunakan untuk membentuk sebuah model *classifier*. *Testing set* digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Memisahkan data menjadi *training* dan *testing* dimaksudkan agar model yang diperoleh memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data [7].

2.3 Klasifikasi (Classification)

Klasifikasi merupakan salah satu peran dalam *data mining*. Klasifikasi bertujuan menentukan suatu model, dimana model ini digunakan untuk prediksi kelompok atau kelas dari data baru. Suatu kumpulan *testing set* digunakan untuk menentukan keakuratan suatu model [8].

2.4 Algoritma Naïve Bayes

Naïve Bayes merupakan teknik prediksi berbasis probabilitas sederhana dengan asumsi semua kategori independen. Kaidah Bayes dikemukakan oleh seorang pendeta Presbyterian Inggris tahun 1763 yang bernama Thomas Bayes. Kaidah ini digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Bentuk umum atau persamaan dari teorema Bayes adalah:

$$P(Y_r|X) = \frac{P(Y_r \cap X)}{\sum_{i=1}^k P(Y_i \cap X)} = \frac{P(Y_r)P(X|Y_r)}{\sum_{i=1}^k P(Y_i)P(X|Y_i)}, \quad r = 1, 2, \dots, k \quad (1)$$

Keterangan [9]:

1. Probabilitas awal (probabilitas *prior*), yaitu probabilitas sebelum ada tambahan informasi, yaitu $P(Y_r)$.
2. Probabilitas bersyarat, yaitu probabilitas dimana terjadinya suatu peristiwa didahului oleh terjadinya peristiwa lain, yaitu $P(X|Y_r)$.
3. Peristiwa ganda, yaitu probabilitas gabungan, yaitu $\{\sum_{i=1}^k P(Y_i) P(X|Y_i)\}$.
4. Probabilitas *posterior*, yaitu probabilitas yang diperbaiki dengan adanya informasi tambahan, yaitu $P(Y_r|X)$.

2.5 Ketepatan Klasifikasi

Pada pengujian ketepatan klasifikasi, akan dihitung ukuran pengamatan data yang diprediksi dengan tingkat kesalahan aktual yang dihasilkan menggunakan APER (*Apparent Error Rate*). APER merupakan ukuran yang menyatakan proporsi sampel data *training* yang salah diklasifikasikan dari model klasifikasi yang telah didapat. APER dapat dihitung menggunakan *confusion matrix*, dimana bentuk *confusion matrix*nya sebagai berikut.

Tabel 1. *Confusion Matrix*

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Total
	Kelas 1	Kelas 2	
Kelas 1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1
Kelas 2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

Keterangan:

- n_{1C} = jumlah amatan kelas 1 yang diklasifikasikan benar sebagai kelas 1
- n_{1M} = jumlah amatan kelas 1 yang diklasifikasikan salah sebagai kelas 2
- n_{2C} = jumlah amatan kelas 2 yang diklasifikasikan benar sebagai kelas 2
- n_{2M} = jumlah amatan kelas 2 yang diklasifikasikan salah sebagai kelas 1

Maka, nilai APER adalah

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{2}$$

Sehingga, untuk menghitung nilai ketepatan klasifikasi yaitu $1 - APER$ [10].

Pengukuran kinerja algoritma klasifikasi dapat dilakukan dengan menggunakan nilai AUC (*Area Under Curve*). Untuk mengetahui nilai AUC adalah sebagai berikut.

$$AUC = \left[\frac{1 + Sensitivity - FPR}{2} \right] \tag{3}$$

Dengan,

$$Sensitivity = \left[\frac{n_{1C}}{n_{1C} + n_{1M}} \right] \tag{4}$$

$$Specificity = \left[\frac{n_{2C}}{n_{2C} + n_{2M}} \right] \tag{5}$$

Dimana,

FPR (False Positive Rate) = $1 - Specificity$. Berikut Tabel kinerja klasifikasi menggunakan nilai AUC [11].

Tabel 2. Nilai AUC (*Area Under Curve*)

Nilai AUC	Keterangan
0,91 – 1,00	Klasifikasi Sangat Baik
0,81 – 0,90	Klasifikasi Baik
0,71 – 0,80	Klasifikasi Cukup
0,61 – 0,70	Klasifikasi Buruk
$\leq 0,60$	Klasifikasi Salah

2.6 Penjualan

Penjualan adalah proses, cara, perbuatan menjual. Menjual merupakan proses memberikan sesuatu kepada orang lain (pembeli) untuk memperoleh uang pembayaran atau menerima uang [12].

2.7 Provider

Provider merupakan penyedia layanan jasa. Dalam suatu komunikasi dibutuhkan *provider* untuk menyambungkan dan membantu dalam meningkatkan produktifitas suatu kegiatan atau pekerjaan, salah satunya yaitu *Internet Service Provider* (ISP). ISP merupakan perusahaan yang menyediakan layanan jasa sambungan internet atau jasa sambungan lainnya [13].

2.8 Pulsa Elektrik

Pulsa elektrik berarti pulsa yang ditransfer melalui aliran listrik. Pulsa adalah satuan unit yang digunakan oleh penyedia layanan telekomunikasi seluler. Dalam dunia telekomunikasi, satuan pulsa ditetapkan sebagai unit satuan dalam menghitung penggunaan layanan telekomunikasi seluler [14].

2.9 Metode Penelitian

Data yang digunakan adalah data penjualan pulsa dari 20 konter yang memiliki pembukuan pulsa di Kecamatan Masbagik, Kabupaten Lombok Timur, NTB. Digunakan teknik *purposive sampling*, dimana *purposive sampling* adalah pengambilan sampel dengan menentukan kriteria sampel yang dibutuhkan [15]. Adapun variabel yang digunakan dalam penelitian ini tercantum pada Tabel 3 berikut:

Tabel 3. Variabel Penelitian

Jenis variabel	Nama Variabel	Kategori	Skala
Variabel independen (X)	Nama <i>Provider</i> (X_1)	1. Axis 2. IM3 3. Smartfren 4. Telkomsel 5. Tri	Nominal

Jenis variabel	Nama Variabel	Kategori	Skala
		6. XL	
	Nominal	1. Rendah	Ordinal
	Pengisian(X_2)	2. Sedang	
		3. Tinggi	
	Laba(X_3)	1. Rendah	Ordinal
		2. Sedang	
		3. Tinggi	
	Peminat(X_4)	1. Sedikit	Ordinal
		2. Sedang	
		3. Banyak	
Variabel dependen (Y)	Status Klasifikasi	1. Tidak Laris	Ordinal
		2. Cukup Laris	
		3. Laris	

Langkah-langkah yang dilakukan pada proses *Naïve Bayes* adalah sebagai berikut :

1. Dilakukan persiapan alat dan bahan untuk menganalisis data penelitian.
2. Dilakukan pengumpulan *dataset*. Jumlah sampel yang digunakan adalah sebanyak 3.084 data dengan menggunakan teknik *purposive sampling* dan *proses cleaning*. Proses *cleaning* berfungsi untuk menghapus data yang tidak lengkap (*missing value*), mengatasi ketidakkonsistenan, dan menghilangkan data yang bersifat *noise*.
3. Dilakukan analisa deskriptif untuk melihat rincian data awal pada variabel penelitian.
4. Dilakukan pembagian *dataset* menjadi data *training* dan data *testing*. Pada penelitian ini, digunakan persentase data *training* 90% dan persentase data *testing* 10%.
5. Dilakukan proses *Naïve Bayes Classification* pada data *training*. Pada proses ini, dilakukan perhitungan probabilitas *prior* dan *posterior*. Dihitung jumlah dan probabilitas setiap kategori dan setiap label untuk memperoleh model terbaik yang digunakan untuk proses klasifikasi data *testing*.
6. Dilakukan proses prediksi menggunakan model yang sudah didapatkan menggunakan data *testing*. Proses ini dilakukan dengan menggunakan *software R* seperti pada persamaan (1).
7. Didapatkan tabel *confusion matrix* dari data *testing* seperti pada Tabel 1. Hasil data *testing* akan diuji ketepatan dan kinerja klasifikasinya menggunakan persamaan (2) dan persamaan (3), sehingga didapatkan kesimpulan.

3. Hasil dan Diskusi

3.1 Statistika Deskriptif

Dari 3.084 data, terdapat 555 data dengan status klasifikasi tidak laris, 1.273 data dengan status klasifikasi cukup laris, dan 1.256 data dengan status klasifikasi laris.

3.2 Pembagian Data

Dari keseluruhan data yang berjumlah 3.084 data, dilakukan pembagian data *training* sebesar 90% (2.775 data) dan data *testing* sebesar 10% (309 data) untuk melakukan pengklasifikasian *Naïve Bayes*.

3.3 *Naïve Bayes Classification* (NBC)

NBC adalah metode klasifikasi yang memanfaatkan perhitungan probabilitas dalam menentukan kelas klasifikasi. Dari 2.775 data *training* yang digunakan, dihasilkan kelas klasifikasi dan tabel prediksi. Berikut langkah-langkah menentukan status klasifikasi dengan menggunakan metode *Naïve Bayes*.

1) Perhitungan Probabilitas *Prior* ($P(Y_r)$)

Pada 2.775 data *training* penjualan *provider* pulsa, terdapat sejumlah pembelian pulsa yang diklasifikasikan ke dalam status klasifikasi laris, status klasifikasi cukup laris, dan status klasifikasi tidak laris. Berikut perhitungan untuk setiap kelas status klasifikasi di data *training* pada Tabel 4 berikut.

Tabel 4. Perhitungan probabilitas *prior* ($P(Y)$)

Status Klasifikasi	Jumlah Pembelian	Probabilitas Status Klasifikasi
Tidak Laris	494	0,178
Cukup Laris	1.155	0,416
Laris	1.126	0,406
Total	2.775	1

2) Probabilitas Kategori Terhadap Masing-masing Kelas ($P(X_i|Y_r)$)

Adapun nilai probabilitas setiap kategori terhadap masing-masing kelas status klasifikasi pada data *training* adalah sebagai berikut.

a. Probabilitas Status Klasifikasi dengan Nama *Provider*

Tabel 5. Probabilitas X_1 bersyarat Y ($P(X_1|Y)$)

Nama <i>Provider</i>	Jumlah Status klasifikasi			Probabilitas Status klasifikasi		
	Tidak Laris	Cukup Laris	Laris	Tidak Laris	Cukup Laris	Laris
Axis	7	33	14	0,014	0,029	0,012
IM3	6	6	10	0,012	0,005	0,009
Smartfren	4	4	0	0,008	0,003	0,000
Telkomsel	202	524	122	0,409	0,454	0,108
Tri	11	6	6	0,022	0,005	0,005
XL	264	582	974	0,534	0,504	0,865
Total	494	1.155	1.126	1	1	1

b. Probabilitas Status Klasifikasi dengan Nominal Pengisian

Tabel 6. Probabilitas X_2 bersyarat Y ($P(X_2|Y)$)

Nominal Pengisian	Jumlah Status klasifikasi			Probabilitas Status klasifikasi		
	Tidak Laris	Cukup Laris	Laris	Tidak Laris	Cukup Laris	Laris
Rendah	63	928	1.126	0,128	0,803	1,000
Sedang	279	227	0	0,565	0,197	0,000
Tinggi	152	0	0	0,308	0,000	0,000
Total	494	1.155	1.126	1	1	1

c. Probabilitas Status Klasifikasi dengan Laba

Tabel 7. Probabilitas X_3 bersyarat Y ($P(X_3|Y)$)

Laba	Jumlah Status klasifikasi			Probabilitas Status klasifikasi		
	Tidak Laris	Cukup Laris	Laris	Tidak Laris	Cukup Laris	Laris
Rendah	152	0	0	0,308	0,000	0,000
Sedang	279	227	0	0,565	0,197	0,000
Tinggi	63	928	1.126	0,128	0,803	1,000
Total	494	1.155	1.126	1	1	1

d. Probabilitas Status Klasifikasi dengan Peminat

Tabel 8. Probabilitas X_4 bersyarat Y ($P(X_4|Y)$)

Peminat	Jumlah Status klasifikasi			Probabilitas Status klasifikasi		
	Tidak Laris	Cukup Laris	Laris	Tidak Laris	Cukup Laris	Laris
Sedikit	28	49	30	0,057	0,042	0,027
Sedang	202	524	122	0,409	0,454	0,108
Banyak	264	582	974	0,534	0,504	0,865
Total	494	1.155	1.126	1	1	1

3.4 Simulasi Perhitungan Data *Testing* dengan Metode *Naïve Bayes Classification* (NBC)

Data *testing* yang digunakan sebanyak 309 data. Dilakukan simulasi perhitungan salah satu data pembelian pulsa ke-35 di data *testing* dengan klasifikasi awal adalah klasifikasi laris, diketahui nama *provider* adalah XL, dengan nominal pengisian pulsa rendah, laba tinggi, dan dengan peminat *provider* XL adalah banyak, maka perhitungan untuk kelas status klasifikasi pada data tersebut berdasarkan hasil di data *training* adalah sebagai berikut.

a. Probabilitas status klasifikasi atau probabilitas *prior* ($P(Y_r)$)

Berdasarkan Tabel 4, probabilitas status klasifikasi tidak laris, status klasifikasi cukup laris, dan status klasifikasi laris berturut-turut adalah 0,178, 0,416, dan 0,406.

b. Probabilitas contoh kasus dengan tiap status klasifikasi

Berdasarkan Tabel 5, probabilitas kategori XL untuk tiap status klasifikasi tidak laris, status klasifikasi cukup laris, dan status klasifikasi laris berturut-turut adalah 0,534, 0,504, dan 0,865.

Berdasarkan Tabel 6, probabilitas nominal pengisian pulsa rendah untuk tiap status klasifikasi tidak laris, status klasifikasi cukup laris, dan status klasifikasi laris berturut-turut adalah 0,128, 0,803, dan 1,000.

Berdasarkan Tabel 7, probabilitas laba tinggi untuk tiap status klasifikasi tidak laris, status klasifikasi cukup laris, dan status klasifikasi laris berturut-turut adalah 0,128, 0,803, dan 1,000.

Berdasarkan Tabel 8, probabilitas peminat *provider* XL banyak untuk tiap status klasifikasi tidak laris, status klasifikasi cukup laris, dan status klasifikasi laris berturut-turut adalah 0,534, 0,504, dan 0,865.

c. Kalikan semua hasil variabel status klasifikasi tidak laris, klasifikasi cukup laris, dan klasifikasi laris ($P(X_i|Y_r)$)

$$P(XL \cap Rendah \cap Tinggi \cap Banyak|Tidak Laris) = 0,005$$

$$P(XL \cap Rendah \cap Tinggi \cap Banyak|Cukup Laris) = 0,164$$

$$P(XL \cap Rendah \cap Tinggi \cap Banyak|Laris) = 0,748$$

d. Perhitungan Probabilitas Gabungan ($\sum_{i=1}^k P(Y_i)P(X|Y_i)$)

$$\begin{aligned} \sum_{i=1}^3 P(Y_i)P(X|Y_i) &= P(Tidak Laris)P(X|Tidak Laris) + \\ &\quad P(Cukup Laris)P(X|Cukup Laris) + P(Laris)P(X|Laris) \\ &= (0,178 \times 0,005) + (0,416 \times 0,164) + (0,406 \times 0,748) \\ &\approx 0,373 \end{aligned}$$

e. Tentukan status klasifikasi dengan teorema *bayes* yaitu probabilitas *posterior* ($P(Y_r|X)$)

$$P(Y_1|X) = \frac{P(Y_1)P(X|Y_1)}{\sum_{i=1}^3 P(Y_i)P(X|Y_i)} = \frac{0,178 \times 0,005}{0,373} = 0,002$$

$$P(Y_2|X) = \frac{P(Y_2)P(X|Y_2)}{\sum_{i=1}^3 P(Y_i)P(X|Y_i)} = \frac{0,416 \times 0,164}{0,373} = 0,183$$

$$P(Y_3|X) = \frac{P(Y_3)P(X|Y_3)}{\sum_{i=1}^3 P(Y_i)P(X|Y_i)} = \frac{0,406 \times 0,748}{0,373} = 0,814$$

Dapat dilihat bahwa nilai probabilitas klasifikasi laris lebih tinggi dibandingkan dengan nilai probabilitas klasifikasi cukup laris dan klasifikasi tidak laris, sehingga data pembelian pulsa ke-35 tetap dalam klasifikasi laris. Hal yang sama dilakukan pada data *testing* lainnya.

3.5 Confusion Matrix Data Testing

Berdasarkan perhitungan probabilitas sebelumnya, dilakukan pengujian data *testing* sebanyak 309 data. Adapun tabel untuk mengevaluasi hasil klasifikasi adalah *confusion*

matrix. Pada Tabel 9 berikut diberikan hasil perhitungan *confusion matrix* untuk data *testing*.

Tabel 9. *Confusion Matrix* Data *Testing*

Aktual	Prediksi			Total
	Tidak Laris	Cukup Laris	Laris	
Tidak Laris	55	4	2	61
Cukup Laris	22	53	43	118
Laris	0	13	117	130
Total	77	70	162	309

Pada Tabel 9 tersebut, dihasilkan 55 data berhasil diklasifikasikan ke tidak laris, 53 data berhasil diklasifikasikan ke cukup laris, dan 117 data berhasil diklasifikasikan ke laris, sehingga dari 309 data *testing* yang digunakan, sebanyak 225 data berhasil diklasifikasikan dengan benar.

3.6 Pengujian Ketepatan Klasifikasi

Pengujian ketepatan klasifikasi pada data *testing* dengan pengklasifikasian *multiclass* berbeda dengan kasus klasifikasi *biner*. Berdasarkan Tabel 9, maka nilai APER dalam *confusion matrix* tersebut adalah:

$$APER = \frac{4 + 2 + 22 + 43 + 0 + 13}{77 + 70 + 162} = 0,272$$

Didapatkan nilai APER 0.272 yang berarti ketepatan klasifikasi yang didapatkan sebesar 0,728 atau 72,8%. Suatu kinerja algoritma klasifikasi dapat dihitung menggunakan nilai AUC pada persamaan (3). Dilakukan perhitungan untuk mencari nilai *sensitivity* dan nilai *specificity* pada Tabel 9 menggunakan persamaan (4) dan persamaan (5). Berdasarkan perhitungan pada *confusion matrix*, didapatkan nilai *sensitivity*, nilai *specificity*, dan nilai FPR berturut-turut sebesar 0,750, 0,857, dan 0,143. Dari hasil tersebut, dilakukan perhitungan nilai AUC (*Area Under Curve*) sebagai berikut:

$$AUC = \frac{1 + Sensitivity - FPR}{2} = \frac{1 + 0,750 - 0,143}{2} = \frac{1,607}{2} = 0,804$$

Berdasarkan Tabel 2 dan nilai AUC yang dihasilkan, dapat disimpulkan bahwa ketepatan klasifikasi penjualan *provider* pulsa sudah cukup.

Kinerja klasifikasi yang dihasilkan tidak begitu baik tapi tidak begitu buruk, sehingga dibutuhkan metode klasifikasi lain untuk menghasilkan kinerja klasifikasi yang lebih baik. Skala data variabel nama *provider* (X_1) berbentuk nominal, dimana data nominal tidak memiliki nilai dan tidak memiliki urutan. Apabila skala data nominal

ditransformasi ke bentuk data numerik (data yang memiliki nilai), maka dapat digunakan metode klasifikasi seperti algoritma KNN, SVM dan lain-lain untuk melihat kinerja klasifikasi yang dihasilkan.

Nilai akurasi yang dihasilkan pada penelitian ini dipengaruhi oleh persentase data *training* dan data *testing*, cara pengklasifikasian, serta objek data yang digunakan. Pemilihan data acak pada data *training* dengan menggunakan fungsi di *software R* menghasilkan beberapa kombinasi model yang menghasilkan nilai akurasi yang berbeda. Hal ini juga berlaku ketika persentase data *training* dan data *testing* yang digunakan berbeda. Dhasilkan nilai akurasi lebih tinggi ketika persentase data *training* lebih tinggi dari persentase data *testing*.

4. Kesimpulan

Berdasarkan hasil yang telah dijelaskan sebelumnya, maka dapat ditarik kesimpulan sebagai berikut:

1. Hasil klasifikasi penjualan *provider* pulsa menggunakan metode *Naïve Bayes* di Kecamatan Masbagik menunjukkan bahwa pada model *Naïve Bayes*, sebanyak 225 data berhasil diklasifikasikan dengan benar dari 309 data *testing* yang digunakan.
2. Ketepatan hasil klasifikasi penjualan *provider* pulsa menggunakan metode *Naïve Bayes* menghasilkan nilai APER 0,272 yang berarti bahwa ketepatan klasifikasi yang didapatkan sebesar 0,728 atau 72,8%. Didapatkan juga nilai AUC sebesar 0,804 yang berarti ketepatan klasifikasi penjualan *provider* pulsa sudah cukup.

Daftar Pustaka

- [1] Hamidin, D., Pranawukir, I., Mulyana, A., Susilawati, E., Ikhrum, F., Novalia, N., Ruminda, M., Dawis, A.M., Kurniawan, R., & Pandriadi. *Strategi Pemasaran Di Era Digital*. Sukabumi: Haura Utama. 2022.
- [2] Palupi, E.S. Prediction Of Android Handphone Sales During Pandemic Using Naïve Bayes and K-NN Methods Based On Particle Swarm Optimization. *Jurnal Riset Informatika*, 23-28. 2021.
- [3] Arifin, T., & Ariesta, D. Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insetif*, 26-30, 2019.
- [4] Nawangsih, I., & Setyaningsih, A. Penerapan Algoritma Naïve Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. *Jurnal Teknologi Pelita Bangsa*, 195-207, 2020.
- [5] Devita, R.N., Herwanto, H.W., & Wibawa, A.P. Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 427-434, 2018.

- [6] Arhami, M., & Nasir, M. *Data Mining Algoritma dan Implementasi*. Ed.1. Yogyakarta: Andi. 2020.
- [7] Prasetyowati, E. *Data Mining: Pengelompokan Data untuk Informasi dan Evaluasi*. Pamekasan: Duta Media. 2017.
- [8] Muflikhah, L., Ratnawati, D.E., & Putri, R.R.M. *Data Mining*. Malang: UB Press. 2018.
- [9] Freund's, J.E., & Walpole, R.E. *Mathematical Statistics with Applications*. 8th Edition. USA: Pearson Education Limited. 2014.
- [10] Johnson, A.R., & Wichern, D. W. *Applied Multivariate Statistical Analysis*. Sixth Edition. Upper Saddle River: New Jersey. 2007.
- [11] Gorunescu, F. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer. 2011.
- [12] Arifin, S. *Sales Management: Strategi Menjual dengan Pendekatan Personal*. Yogyakarta: Salma Idea. 2020.
- [13] Putri, M.P., Budiman, E., & Taruk, M. Analisis Kualitas Jaringan Seluler Terhadap Jasa Provider di Kota Samarinda. *Jurnal Politeknik Negeri Balikpapan*, 322-325, 2017.
- [14] Priyantomo, B. *Panduan Startup Server Pulsa: Panduan Untuk Anda Yang Ingin Berbisnis Server Pulsa*. Malaysia: Mobile Outlet, 2016.
- [15] Sugiyono. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta. 2012.