

KLASIFIKASI PENJUALAN PROVIDER PULSA DI KECAMATAN MASBAGIK LOMBOK TIMUR MENGGUNAKAN METODE NAÏVE BAYES

CREDIT PROVIDER SALES CLASSIFICATION IN MASBAGIK DISTRICT EAST LOMBOK USING THE NAÏVE BAYES METHOD

NURUL MUSFITA^{1,*,}, NURUL FITRIYANI^{2,}, ZULHAN WIDYA BASKARA³

^{1,2,3} Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram.
Jl. Majapahit No.62, Mataram 83115, Nusa Tenggara Barat, Indonesia. Tel./Fax. (0370) 633007,

*email: nurul.musfita18@gmail.com.

Abstrak. Perkembangan teknologi yang pesat menyebabkan proses komunikasi dan informasi menggunakan *handphone* meningkat. *Handphone* membutuhkan jaringan internet untuk berkomunikasi, sehingga kebutuhan pulsa elektrik untuk mengakses komunikasi dan informasi juga 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* atau penyedia layanan komunikasi di wilayah tersebut. Untuk melihat jenis *provider* yang paling banyak digunakan di Kecamatan Masbagik Lombok Timur dengan mempertimbangkan jumlah penduduk terbanyak, dilakukan pengklasifikasian untuk memprediksi apakah *provider* tersebut laris atau tidak menggunakan metode *Naïve Bayes*. Tujuan penelitian ini adalah untuk menentukan hasil klasifikasi dan ketepatan hasil 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 bahwa ketepatan klasifikasi penjualan *provider* pulsa yang laris, cukup laris, dan tidak laris sudah cukup.

Kata kunci: Akurasi, Klasifikasi, *Naïve Bayes Classification* (NBC), *Provider*.

Abstract. The rapid development of technology causes the process of communication and information using mobile phones to increase. Mobile phones require an internet network to communicate, so the need for electrical pulses to access communication and information also increases. East Lombok is the area with the largest population in NTB and the high users of information technology. East Lombok also has an internet network or smooth communication signal, which shows that there are many providers of communication service in the area. To see the types of providers that are most widely used in Masbagik District, East Lombok, taking into account the largest population, a classification is carried out to predict whether these providers are in demand or not using the *Naïve Bayes* method. The purpose of this study was to determine the classification results and the accuracy of the results 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 is sufficient.

Keywords: Accuracy, Classification, *Naïve Bayes Classification* (NBC), *Provider*.

PENDAHULUAN

Manusia merupakan makhluk sosial yang memiliki kebutuhan dasar dalam bersosialisasi, dalam proses sosialisasi membutuhkan proses komunikasi (Sari, 2017). 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 (Hamidin *et al.*, 2022).

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, sudah 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, seperti Telkomsel/AS, XL, Axis, Smartfren, Tri, Indosat/IM3 dan lainnya. 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 dengan menggunakan pulsa dan kuota. Hal ini menyebabkan kebutuhan dan penggunaan pulsa elektrik semakin meningkat (Palupi, 2021). Salah satu pertimbangan pengguna jasa telekomunikasi adalah kecepatan internet dari suatu *provider* seluler. 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 guna melakukan inferensi induksi pada persoalan klasifikasi (Annur, 2018). *Naïve Bayes* memiliki beberapa kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana, serta berakurasi tinggi (Arifin dan Ariesta, 2019).

Digunakan *software R* untuk pengklasifikasiannya karena bersifat *open source*, yaitu bahasa *R* dapat digunakan untuk berbagai kepentingan baik komersial maupun nonkomersial tanpa perlu lisensi atau biaya, termasuk untuk melakukan operasi *machine learning* seperti klasifikasi dan regresi.

Penelitian ini didasarkan pada penelitian sebelumnya oleh Putro *et al.* (2020) yang berjudul “Penerapan Metode *Naive Bayes* Untuk Klasifikasi Pelanggan”. Hasil yang diperoleh dalam mengklasifikasikan pelanggan berpotensi dan tidak berpotensi menunjukkan nilai *Accuracy* 92%, nilai *Precision* 100%, dan nilai *Recall* 91%. Adapun penelitian oleh Nawangsih dan Setyaningsih (2020) tentang Penerapan Algoritma *Naive Bayes* Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa, menunjukkan bahwa hasil klasifikasi dan akurasi data terhadap stok penjualan pulsa menghasilkan nilai *Accuracy* 97,50%, nilai *Precision* 100%, dan nilai *Recall* 93,48%. Pada penelitian Devita *et al.* (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 *Naive Bayes* memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 40%.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menentukan hasil klasifikasi dan ketepatan hasil klasifikasi penjualan *provider* pulsa menggunakan metode *Naive 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.

MATERI DAN METODE

1. MATERI

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-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan (Arhami dan Nasir, 2020).

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 set* dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data (Prasetyowati, 2017).

Klasifikasi (Classification)

Klasifikasi merupakan peran dalam *data mining* yang menggunakan metode pendekatan prediktif. Klasifikasi bertujuan menentukan suatu model, dimana model ini digunakan untuk prediksi kelompok atau kelas dari data baru. Suatu kumpulan data uji (*testing set*) digunakan untuk menentukan keakuratan suatu model (Muflikhah *et al.*, 2018).

Algoritma Naïve Bayes

Naïve Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan Bayes dengan asumsi ketidaktergantungan yang kuat yaitu kemunculan fitur tertentu tidak tergantung pada kemunculan fitur lain. *Naïve Bayes* merupakan salah satu metode yang efektif dan juga efisien pada proses pengklasifikasian (Prasetyo, 2012). Kaidah Bayes atau Teori 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 (Supranto, 1985). Bentuk umum atau persamaan dari teorema Bayes adalah (Freund and Walpole, 2014):

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

Keterangan:

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 $P(Y_r|X)$.

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

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \quad (2)$$

Sehingga, untuk menghitung nilai ketepatan klasifikasi adalah $1 - APER$ (Johnson and Wichern, 2007).

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

$$AUC = \left[\frac{1 + \text{Sensitivity} - \text{FPR}}{2} \right] \quad (3)$$

Dengan,

$$\text{Sensitivity} = \left[\frac{n_{1C}}{n_{1C} + n_{1M}} \right] \quad (4)$$

$$\text{Specificity} = \left[\frac{n_{2C}}{n_{2C} + n_{2M}} \right] \quad (5)$$

Dimana,

FPR (*False Positive Rate*) = $1 - \text{Specificity}$. Berikut Tabel kinerja klasifikasi menggunakan nilai AUC.

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
≤ 0.60	Klasifikasi Salah

(Sumber: Gorunescu, 2011)

Penjualan

Penjualan adalah proses, cara, perbuatan menjual (KBBI, 2022). Menjual merupakan proses memberikan sesuatu kepada orang lain (pembeli) untuk memperoleh uang pembayaran atau menerima uang (Arifin, 2020).

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)*. *Internet Service Provider* merupakan suatu perusahaan yang menyediakan layanan jasa sambungan internet atau jasa sambungan lainnya (Putri *et al.*, 2017).

Pulsa Elektrik

Kata “elektrik” dalam KBBI memiliki makna “listrik”. Oleh karena itu, pulsa elektrik berarti pulsa yang ditransfer melalui aliran listrik (Inayati, 2017). Pulsa adalah sebuah satuan unit yang digunakan oleh penyedia layanan telekomunikasi seluler. Dalam dunia telekomunikasi, satuan pulsa ditetapkan sebagai unit satuan untuk mempermudah dalam menghitung penggunaan layanan telekomunikasi seluler (Priyantomo, 2016).

2. METODE

Lokasi Penelitian

Lokasi penelitian ini adalah Kecamatan Masbagik, Kabupaten Lombok Timur, NTB.

Prosedur Kerja

Data yang digunakan adalah data penjualan pulsa. Data dan informasi dikumpulkan dengan melakukan observasi, wawancara, dan studi pustaka terkait penelitian sebelumnya. Digunakan teknik *purposive sampling*, dimana *purposive sampling* adalah pengambilan sampel dengan menentukan kriteria sampel yang dibutuhkan atau dengan pertimbangan tertentu (Sugiyono, 2012). 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 Provider(X_1)	1. Axis	Nominal
		2. IM3	
		3. Smartfren	
		4. Telkomsel	
Variabel independen (X)	Pengisian(X_2)	5. Tri	Ordinal
		6. XL	
		1. Rendah	
Variabel independen (X)	Laba(X_3)	2. Sedang	Ordinal
		3. Tinggi	
		1. Rendah	
Variabel independen (X)	Peminat(X_4)	2. Sedang	Ordinal
		3. Banyak	
		1. Sedikit	
Variabel dependen (Y)	Status Klasifikasi	2. Cukup Laris	Ordinal
		3. Laris	
		1. Tidak Laris	

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

1. Dilakukan persiapan alat dan bahan yang digunakan untuk menganalisis data penelitian.
2. Dilakukan pengumpulan *dataset*. Jumlah sampel yang digunakan adalah sebanyak 3084 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 visualisasi data awal pada variabel penelitian. Visualisasi data menggunakan tabel, gambar, atau grafik.
4. Dilakukan pembagian *dataset* menjadi data *training* dan data *testing*. Pembagian data didasarkan pada pengklasifikasian penelitian sebelumnya, yaitu diantara 60%-90% untuk data *training* dan sisanya untuk data *testing* (Han *et al.*, 2006). Pada penelitian ini, digunakan persentase data *training* 90% yaitu 2775 data dan persentase *testing* 10% yaitu 309 data.
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, sehingga didapatkan kesimpulan.

Analisis Data

Alat yang digunakan dalam penelitian ini adalah laptop, *Microsoft Excel* dan *software* R versi 4.0.5. *Dataset* yang digunakan merupakan data penjualan pulsa pada tanggal 25-31 Agustus 2022 di 20 konter yang memiliki pembukuan pulsa di Kecamatan Masbagik, Kabupaten Lombok Timur. Dilakukan analisis deskriptif dan pengklasifikasian data pada sampel berjumlah 3084 menggunakan distribusi frekuensi. Data tersebut digunakan untuk memprediksi *provider* terlaris dalam penjualan pulsa dengan menjadikan *training* 90% dan *testing* 10%. Selanjutnya dilakukan proses *Naïve Bayes Classification* (NBC) pada data *training* menggunakan persamaan (1) sehingga didapatkan model terbaik dari data *training*. Hal yang sama dilakukan pada data *testing* hingga didapatkan hasil *confusion matrix* data *testing* seperti pada Tabel 1. Hasil klasifikasi data *testing* akan diuji ketepatan pengklasifikasiannya menggunakan persamaan (2) dan persamaan (3).

HASIL DAN DISKUSI

Statistika Deskriptif

Dari 3084 data, terdapat 555 data dengan status klasifikasi tidak laris, 1273 data dengan status klasifikasi cukup laris, dan 1256 data dengan status klasifikasi laris.

Pembagian Data

Dari keseluruhan data yang berjumlah 3084 data, dilakukan pembagian data *training* sebesar 90% (2775 data) dan data *testing* sebesar 10% (309 data) untuk melakukan pengklasifikasian *Naïve Bayes*.

Naïve Bayes Classification (NBC)

Naïve Bayes Classification (NBC) adalah metode klasifikasi yang memanfaatkan perhitungan probabilitas dalam menentukan kelas klasifikasi. Dari 2775 data *training* yang digunakan, dihasilkan kelas klasifikasi dan tabel prediksi. Berikut langkah-langkah dalam menentukan status klasifikasi dengan menggunakan metode *Naïve Bayes*.

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

Pada 2775 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	1155	0.416
Laris	1126	0.406
Total	2775	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*

Berikut perhitungan probabilitas untuk setiap status klasifikasi penjualan dengan nama *provider* pada Tabel 5 berikut.

Tabel 5. Probabilitas Status Klasifikasi dengan Variabel Nama *Provider*

Nama Provider	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	1155	1126	1	1	1

b. Probabilitas Status Klasifikasi dengan Nominal Pengisian

Berikut perhitungan probabilitas untuk setiap status klasifikasi penjualan dengan nominal pengisian pada Tabel 6 berikut.

Tabel 6. Probabilitas Status Klasifikasi dengan Variabel Nominal Pengisian

Nominal Pengisian	Jumlah Status klasifikasi			Probabilitas Status klasifikasi		
	Tidak Laris	Cukup Laris	Laris	Tidak Laris	Cukup Laris	Laris
Rendah	63	928	1126	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	1155	1126	1	1	1

c. Probabilitas Status Klasifikasi dengan Laba

Berikut perhitungan probabilitas untuk setiap status klasifikasi penjualan dengan laba pada Tabel 7 berikut.

Tabel 7. Probabilitas Status Klasifikasi dengan Variabel Laba

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	1126	0.128	0.803	1.000
Total	494	1155	1126	1	1	1

d. Probabilitas Status Klasifikasi dengan Peminat

Berikut perhitungan probabilitas untuk setiap status klasifikasi penjualan dengan peminat pada Tabel 8 berikut.

Tabel 8. Probabilitas Status Klasifikasi dengan Variabel Peminat

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	1155	1126	1	1	1

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

Data *testing* merupakan data yang digunakan untuk menguji model yang didapatkan. 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 adalah sebagai berikut.

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

Berdasarkan Tabel 4, probabilitas tiap status klasifikasi ($P(Y_r)$) yaitu $P(\text{Tidak Laris}) = 0.178$, $P(\text{Cukup Laris}) = 0.416$, dan $P(\text{Laris}) = 0.406$.

b. Probabilitas contoh kasus dengan tiap status klasifikasi

Berdasarkan Tabel 5, probabilitas XL untuk tiap status klasifikasi yaitu $P(\text{XL}|\text{Tidak Laris}) = 0.534$, $P(\text{XL}|\text{Cukup Laris}) = 0.504$, dan $P(\text{XL}|\text{Laris}) = 0.865$.

Berdasarkan Tabel 6, probabilitas nominal pengisian pulsa rendah untuk tiap status klasifikasi yaitu $P(\text{Rendah}|\text{Tidak Laris}) = 0.128$, $P(\text{Rendah}|\text{Cukup Laris}) = 0.803$, dan $P(\text{Rendah}|\text{Laris}) = 1.000$.

Berdasarkan Tabel 7, probabilitas laba tinggi untuk tiap status klasifikasi yaitu $P(\text{Tinggi}|\text{Tidak Laris}) = 0.128$, $P(\text{Tinggi}|\text{Cukup Laris}) = 0.803$, dan $P(\text{Tinggi}|\text{Laris}) = 1.000$.

Berdasarkan Tabel 8, probabilitas peminat *provider* XL banyak untuk tiap status klasifikasi yaitu $P(\text{Banyak}|\text{Tidak Laris}) = 0.534$, $P(\text{Banyak}|\text{Cukup Laris}) = 0.504$, dan $P(\text{Banyak}|\text{Laris}) = 0.865$.

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

- $P(\text{XL} \cap \text{Rendah} \cap \text{Tinggi} \cap \text{Banyak}|\text{Tidak Laris}) = 0.534 \times 0.128 \times 0.128 \times 0.534 = 0.005$

- $P(XL \cap Rendah \cap Tinggi \cap Banyak | Cukup Laris) = 0.504 \times 0.803 \times 0.803 \times 0.504 = 0.164$
- $P(XL \cap Rendah \cap Tinggi \cap Banyak | Laris) = 0.865 \times 1.000 \times 1.000 \times 0.865 = 0.748$

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

$$\begin{aligned} \sum_{i=1}^3 P(Y_i)P(X|Y_i) &= P(\text{Tidak Laris})P(X|\text{Tidak Laris}) + \\ &\quad P(\text{Cukup Laris})P(X|\text{Cukup Laris}) + P(\text{Laris})P(X|\text{Laris}) \\ &= (0.178 \times 0.005) + (0.416 \times 0.164) + (0.406 \times 0.748) \\ &\approx 0.373 \end{aligned}$$

e. Tentukan status klasifikasi dengan teorema *bayes* atau 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.

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.

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 bahwa ketepatan klasifikasi yang didapatkan sebesar 0.728 atau 72.8%. Mengukur 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 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.

Diskusi

Kinerja klasifikasi yang dihasilkan tidak begitu baik tapi tidak begitu buruk, sehingga dibutuhkan metode klasifikasi yang lain untuk mendapatkan kinerja klasifikasi yang lebih baik. Pada variabel nama *provider* (X_1), skala datanya berbentuk nominal, dimana data nominal tidak memiliki nilai dan tidak memiliki urutan. Apabila skala data nominal pada variabel X_1 ditransformasi ke bentuk data numerik (data yang memiliki nilai), maka dapat digunakan metode klasifikasi seperti algoritma *K-Nearest Neighbor*, *Support Vector Machine* (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, dan 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. Dihasilkan nilai akurasi lebih tinggi ketika persentase data *training* lebih tinggi dari persentase data *testing*.

KESIMPULAN

Berdasarkan rumusan masalah 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

- Annur, H. 2018. Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes. *ILKOM Jurnal Ilmiah* 10(2): 160–165.
- Arhami, M. dan Nasir, M. 2020. *Data Mining Algoritma dan Implementasi*. Ed.1. Yogyakarta: Andi.
- Arifin, S. 2020. *Sales Management: Strategi Menjual dengan Pendekatan Personal*. Yogyakarta: Salma Idea.
- Arifin, T. dan Ariesta, D. 2019. Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insetif* 13(1): 26-30.
- Devita, R.N., Herwanto, H.W., dan Wibawa, A.P. 2018. Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* 5(4): 427-434.
- Freund's, J.E. and Walpole, R.E. 2014. *Mathematical Statistics with Applications*. 8th Edition. USA: Pearson Education Limited.
- Gorunescu, F. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Hamidin, D., Pranawukir, I., Mulyana, A., Susilawati, E., Ikham, F., Novalia, N., Ruminda, M., Dawis, A.M., Kurniawan, R., dan Pandriadi. 2022. *Strategi Pemasaran Di Era Digital*. Sukabumi: Haura Utama.
- Han, J., Kamber, M., and Pei, J. 2006. *Data Mining: Concepts and Techniques*. 2nd Edition. Waltham: Morgan Kaufmann Publishers.
- Inayati, I. 2017. Pulsa Elektronik atau Pulsa Elektrik? (<https://kantorbahasababel.kemdikbud.go.id/info/info-bahasa/pulsa-elektronik-atau-pulsa-elektrik/>). Diunduh jam 23.13 WITA. Tanggal 24 April 2022.
- Johnson, A.R. and Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. Upper Saddle River: New Jersey.
- Muflikhah, L., Ratnawati, D.E., dan Putri, R.R.M. 2018. *Data Mining*. Malang: UB Press.
- Nawangsih, I. dan Setyaningsih, A. 2020. Penerapan Algoritma Naïve Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. *Jurnal Teknologi Pelita Bangsa* 10(4): 195-207.

- Palupi, E.S. 2021. Prediction Of Android Handphone Sales During Pandemic Using Naïve Bayes and K-NN Methods Based On Particle Swarm Optimization. *Jurnal Riset Informatika* 4(1): 23-28.
- Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Prasetyowati, E. 2017. *Data Mining: Pengelompokan Data untuk Informasi dan Evaluasi*. Pamekasan: Duta Media.
- Priyantomo, B. 2016. *Panduan Startup Server Pulsa: Panduan Untuk Anda Yang Ingin Berbisnis Server Pulsa*. Malaysia: Mobile Outlet.
- Putri, M.P., Budiman, E., dan Taruk, M. 2017. Analisis Kualitas Jaringan Seluler Terhadap Jasa Provider di Kota Samarinda. *Jurnal Politeknik Negeri Balikpapan* 2(1): 322-325.
- Putro, H.F., Vlandari, R.T., dan Saptomo, W.L. 2020. Penerapan Metode Naïve Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi* 8(2): 19-24.
- Sari, A.A. 2017. *Komunikasi Antarpribadi*. Yogyakarta: Deepublish.
- Sugiyono. 2012. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Supranto, J. 1985. *Pengantar Probabilitas dan Statistka Induktif*. Jilid 1. Jakarta: Erlangga.