

ANALISIS ALGORITMA NAIVE BAYES CLASSIFIER (NBC) PADA KLASIFIKASI TINGKAT MINAT BARANG DI TOKO VIOLET CELL

Alfan Zainal Macfud, Abdi Pandu Kusuma, Wahyu Dwi Puspitasari

Program Studi Teknik Informatika S1, Fakultas Teknologi Informasi, Universitas Islam Balitar Blitar
Jl. Majapahit No.2-4, Sananwetan, Kec. Sananwetan, Kota Blitar, Jawa Timur 66137
alfan.zainal123@gmail.com

ABSTRAK

Naïve Bayes Classifier yaitu salah satu metode *machine learning* yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memperkirakan probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Algoritma *Naïve Bayes Classifier* dapat diterapkan pada klasifikasi tingkat minat suatu barang untuk menentukan strategi pemasaran. Toko Violet Cell merupakan toko yang menjual berbagai jenis paket kuota internet dengan bermacam merk dengan sistem pencatatan penjualan manual yang kesulitan untuk melakukan klasifikasi terhadap tingkat minat penjualan suatu barang. Tujuan utama penelitian ini yakni untuk mengetahui bagaimana klasifikasi tingkat minat paket kuota internet pada Toko Violet Cell. Penelitian ini menggunakan metode penelitian deskriptif kuantitatif. Data yang digunakan adalah data sekunder berupa data penjualan dalam kurun waktu 2 bulan dan proses klasifikasi minat pada data menggunakan metode *Naïve Bayes Classifier*. Hasil klasifikasi tingkat minat barang menggunakan algoritma *Naïve Bayes Classifier* adalah 23 barang memiliki kategori “Minat Rendah”, 5 barang dikategorikan “Minat Sedang”, dan 1 barang memiliki kategori “Minat Tinggi”. Berdasarkan perhitungan Confusion Matrix hasil klasifikasi yang memakai algoritma *Naïve Bayes Classifier* memperoleh tingkat akurasi sebesar 82,76%.

Kata kunci: *Naïve bayes classifier, tingkat minat barang, penjualan*

1. PENDAHULUAN

Data mining atau penggalian data pada era digital telah menjadi langkah strategis untuk digunakan pada berbagai bidang. Salah satu bidang yang menggunakan *data mining* adalah bidang pemasaran. Di bidang pemasaran, *data mining* dapat memberikan gambaran tingkat minat suatu barang dan membantu penyusunan strategi pemasaran. Selain itu, keterbatasan kemampuan manusia untuk melakukan perhitungan dan pengolahan data secara manual tanpa bantuan komputer membuat perolehan informasi juga terbatas. Oleh karena itu, penerapan *data mining* untuk klasifikasi tingkat minat dari suatu barang sangat diperlukan.

Toko Violet Cell merupakan toko yang menjual berbagai jenis paket kuota internet dengan bermacam merk. Toko tersebut menggunakan sistem pencatatan penjualan manual sehingga dengan sistem manual tersebut pemilik toko kesulitan untuk melakukan klasifikasi terhadap tingkat minat penjualan suatu barang. Oleh karena itu, diperlukan suatu algoritma yang dapat melakukan klasifikasi tingkat minat pembeli pada suatu barang.

Algoritma *Naïve Bayes Classifier* (NBC) yaitu salah satu algoritma teknik klasifikasi. *Naïve Bayes* merupakan algoritma pengklasifikasi dengan metode probabilitas dan statistik yang diutarakan oleh ilmuwan Inggris Thomas Bayes. Algoritma ini berjalan dengan cara memperkirakan peluang di masa depan didasari pengalaman dimasa sebelumnya. Kelebihan algoritma ini adalah diperlukannya jumlah data latih yang kecil dalam menentukan parameter yang dibutuhkan pada proses klasifikasi. Berdasarkan

pemaparan tersebut, algoritma *Naïve Bayes Classifier* dirasa sesuai untuk diterapkan pada proses analisis tingkat minat barang pada Toko Violet Cell.

2. TINJAUAN PUSTAKA

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 [9].

Data mining merupakan proses *iterative* dan interaktif untuk memperoleh pola atau model baru yang sempurna, memiliki manfaat dan dapat dimengerti pada suatu *database* yang sangat besar (*massive database*). *Data mining* berisi pencarian *trend* atau pola yang dibutuhkan dalam database besar untuk membantu pengambil keputusan pada waktu yang akan datang, perangkat tertentu mengenali pola-pola ini akan dapat memberikan suatu analisa data yang bermanfaat dan berwawasan selanjutnya dapat dipelajari dengan lebih teliti, yang dimungkinkan memakai perangkat pendukung keputusan yang lain [1].

Berdasarkan pemaparan tersebut, dapat disimpulkan bahwa *data mining* merupakan proses yang interaktif dan *iterative* dengan menggunakan multidisiplin ilmu untuk mengolah database besar dengan data mentah untuk memperoleh analisis data serta informasi yang bermanfaat dan dapat dimengerti oleh manusia secara awam.

2.2. Naive Bayes Classifier

Data mining mempunyai salah satu teknik yaitu teknik klasifikasi, yang berguna untuk memprediksi suatu nilai dari target variabel kategori [8]. Algoritma Naive Bayes yakni salah satu algoritma teknik klasifikasi. Naive Bayes adalah pengategorian dengan metode probabilitas dan statistik yang diutarakan oleh ilmuwan Inggris Thomas Bayes, yaitu memperkirakan peluang di masa depan didasari pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive yang mana diasumsikan keadaan antar atribut saling bebas [2]. Jadi didalam penerapannya, algoritma *Naive Bayes Classifier* tidak ada hubungan antara satu atribut dengan atribut lain, atau dengan kata lain satu atribut tidak berpengaruh dengan atribut yang lain, sekalipun mungkin atribut tersebut saling berhubungan.

Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis training data. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $Y=F(X)$ dimana Y adalah kelas hasil prediksi dan X adalah tuple yang ingin diprediksi kelasnya [10].

Berikut ini alur dari metode Naive Bayes Classifier [2]:

1. Menghitung nilai peluang kondisi baru (X_k) dari setiap hipotesa terhadap kelas (C_i) yang ada.
2. Menghitung nilai akumulasi peluang dari setiap kelas (C_i)
3. Menghitung Nilai $P(X|C_i) \times P(C_i)$
4. Menentukan kelas dari kasus baru tersebut.

Diungkapkan oleh Siregar dkk (2020), kelemahan metode Naive Bayes ini sendiri yaitu adanya asumsi atau dengan kata lain kondisi kelas saling bebas, sehingga kurang akurat [12], sedangkan pada prakteknya, beberapa kondisi biasanya saling berpengaruh satu sama lain.

2.3. Probabilitas

Probabilitas berasal dari kata probably, yaitu kemungkinan. Jadi probabilitas artinya kemungkinan yang bisa terjadi pada suatu peristiwa [11].

Berdasarkan Ota [3] berpengaruh atau tidaknya suatu probabilitas atas kejadian satu terhadap kejadian yang lain dibagi menjadi dua, yaitu:

1. Probabilitas Dependen
Dua kejadian disebut dependen apabila terjadi atau tidaknya suatu kejadian “berpengaruh” pada probabilitas yang kejadian yang lain.
2. Probabilitas Independen
Dua kejadian disebut independen apabila terjadi atau tidaknya suatu kejadian “tidak berpengaruh” pada probabilitas kejadian yang lain.

Dari dua jenis probabilitas diatas, *Naive Bayes Classifier* merupakan jenis probabilitas independen karena satu kejadian atau satu atribut, tidak berpengaruh pada kejadian atau atribut lain.

2.4. Supervised Learning

Supervised learning merupakan salah satu teknik *machine learning* yang menggunakan dataset (data *training*) yang sudah berlabel (*labeled data*) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi, sedangkan *unsupervised learning* adalah teknik dengan menarik kesimpulan berdasarkan dataset yang merupakan *input data labeled response* [4].

Begitu pula pada penelitian ini, pada *Naive Bayes Classifier* data juga diberi label sedemikian rupa dengan syarat yang telah ditentukan sebelumnya.

2.5. Data Preprocessing

Preprocessing data adalah hal yang harus dilakukan dalam proses *data mining*, karena tidak semua data atau atribut data dalam data digunakan dalam proses *data mining*. Proses ini dilakukan agar data yang akan digunakan sesuai kebutuhan [5].

Data mentah yang di dapat dari sumber luar seringkali memiliki format, bentuk atau pola yang tidak teratur atau tidak normal. Data yang seperti ini harus diproses terlebih dahulu sebelum digunakan, hal ini dilakukan agar proses perhitungan dapat berjalan dengan baik serta tidak terdapat *error* maupun bias karena terdapat kesalahan data didalamnya.

2.6. Aplikasi RapidMiner

RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Aplikasi ini umumnya digunakan dalam bidang yang melibatkan jumlah data yang besar seperti analisa penjualan, penelitian, pendidikan, dan pembelajaran.

3. METODE PENELITIAN

3.1. Jenis Data

Pada penelitian ini data yang digunakan termasuk data sekunder karena merupakan data penjualan yang diperoleh dari arsip penjualan yang ditulis secara manual.

3.2. Pengumpulan Data

Tahap pertama pada penelitian ini yaitu pengumpulan data. Data penjualan selama dua bulan (Mei 2022 – Juni 2022) nantinya akan diolah untuk dijadikan *dataset* klasifikasi tingkat minat paket kuota internet. Produk paket internet yang akan diolah pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Daftar Paket Internet

No.	Provider	Unlimited		Jangka Aktif
		Ya	Tidak	
1.	XL		√	1 Bulan
2.	AXIS		√	1 Bulan
3.	INDOSAT	√	√	5 Hari – 1 Bulan
4.	TRI		√	1 Bulan
5.	TELKOMSEL		√	3 Hari – 1 Bulan
6.	SMARTFREN	√	√	28 Hari – 1 Bulan

3.3. Konversi Data

Data penjualan selama dua bulan tersebut kemudian dikonversi ke dalam bentuk digital berupa *spreadsheet*. Setelah itu, data diolah pada tahap *preprocessing* berupa *labelling* dan *cleaning*. Pada proses ini, data akan diklasifikasikan besaran kuota, kategori harga, dan tingkat penjualannya

Tabel 2. Konversi Data Penjualan

Kategori	Rendah	Sedang	Tinggi
Kuota	Kuota $\leq$ 5 Gb	5 Gb < Kuota $\leq$ 10 Gb	Kuota > 10 Gb
Harga	Harga $\leq$ Rp 25.000	25.000 < Harga $\leq$ Rp 25.001	Harga > Rp 50.000
Jumlah Terjual	Jumlah $\leq$ 5 buah	5 < Jumlah $\leq$ 10 buah	Jumlah > 10 buah

Dataset yang telah diberi label-label tersebut kemudian dilakukan *cleaning* dengan menghapus kolom yang berisi teks yang tidak diperlukan pada proses klasifikasi.

3.4. Data Cleaning

Data *cleaning* atau pembersihan data merupakan proses yang digunakan untuk mendeteksi, memperbaiki ataupun menghapus dataset, tabel, dan database yang korup atau tidak akurat [7].

Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi) [6].

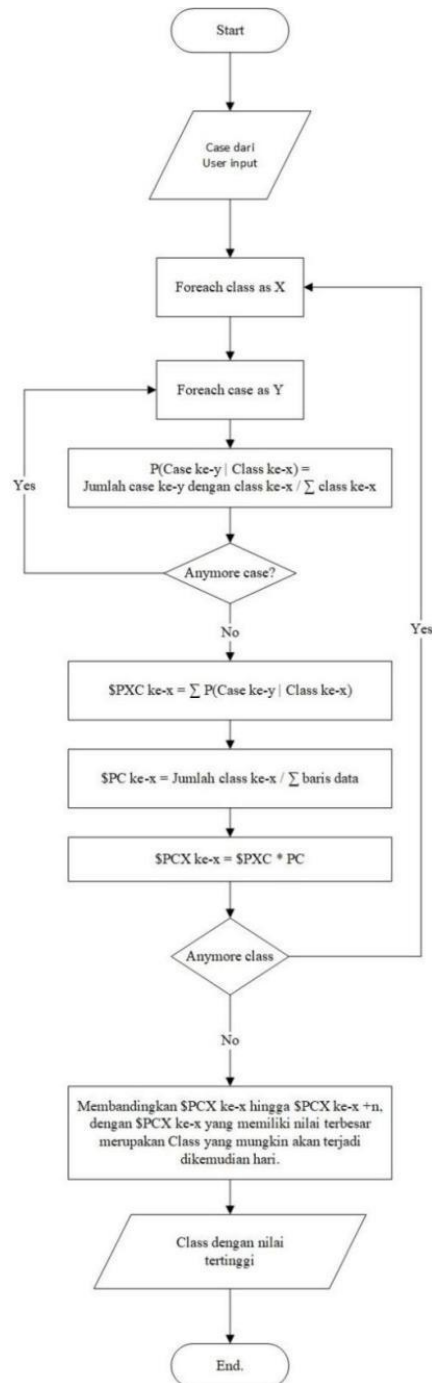
3.5. Klasifikasi

Data mining mempunyai salah satu teknik yaitu teknik klasifikasi, yang berguna untuk memprediksi suatu nilai dari target variabel kategori [8].

3.6. Penerapan Algoritma Naive Bayes

Proses penerapan algoritma pada aplikasi membutuhkan *flowchart* yang mewakili beberapa proses dari algoritma *Naive Bayes*. *Flowchart* ini dapat digunakan untuk mempermudah peneliti untuk menentukan seberapa benar atau seberapa tepat kode program telah ditulis dan seberapa tepat hasil yang telah didapat dari eksekusi kode program tersebut.

Pada gambar 1, proses pertama, program akan menghitung jumlah *case* dengan class tertentu dibagi dengan jumlah kemunculan *class* tersebut pada keseluruhan data. Proses selanjutnya yaitu mengalikan semua hasil perhitungan dari proses 1. Setelah itu, program akan menghitung kemunculan *class* tertentu dibagi jumlah keseluruhan data. Proses keempat adalah dilakukan perkalian hasil dari proses kedua dan ketiga. Proses terakhir yaitu membandingkan hasil proses keempat dari masing-masing *class*. Perhitungan dengan nilai terbesar menentukan *class* mana *case* tersebut akan ditempatkan.



Gambar 1. Flowchart Klasifikasi dengan NBC

Tahap selanjutnya yaitu penerapan algoritma *Naive Bayes Classifier* (NBC) untuk klasifikasi tingkat minat paket kuota internet pada Toko Violet Cell berdasarkan dataset yang telah diolah pada *preprocessing*. Pada sebuah dataset, setiap baris/dokumen diasumsikan sebagai *vector* dari nilai-nilai atribut $\langle x_1, x_2, \dots, x_n \rangle$ dimana tiap nilai-nilai menjadi peninjauan atribut $X_i(i \in [1, n])$. Setiap baris mempunyai label kelas $c_i \in \{c_1, c_2, \dots, c_k\}$ sebagai nilai variabel kelas C , sehingga untuk melakukan klasifikasi dapat dihitung nilai probabilitas $p(C=c_i | X=x_j)$.

Tabel 3. Sampel Dataset Penjualan

No.	Kuota	Kategori harga	Unlimited	Jangka aktif	Provider	Minat
1	Besar	Mahal	Tidak	1 Bulan	XL	Rendah
2	Kecil	Murah	Tidak	1 Bulan	XL	Tinggi
3	Besar	Sedang	Tidak	1 Bulan	XL	Sedang
4	Besar	Sedang	Tidak	1 Bulan	XL	Rendah
5	Kecil	Murah	Tidak	1 Bulan	AXIS	Sedang
6	Sedang	Murah	Tidak	1 Bulan	AXIS	Rendah
7	Kecil	Murah	Tidak	1 Bulan	AXIS	Sedang

Berdasarkan *dataset* tersebut, peneliti ingin mengetahui klasifikasi 3 *class* yang akan muncul yaitu Minat Rendah, Minat Tinggi, dan Minat Sedang dengan case sebagai berikut.

1. Kuota = Besar
2. Kategori Harga = Sedang
3. *Unlimited* = Ya
4. Jangka aktif = 1 Bulan
5. *Provider* = XL

Berikut ini contoh perhitungan klasifikasi setiap *class* menggunakan *Naïve Bayes Classifier*.

a. *Class* 1: Minat Rendah

1. Hitung peluang setiap atribut *case* terhadap *Class* 1
 - a) $P(\text{Kuota Besar}|C1) = 3/7 = 0.43$
 - b) $P(\text{Harga Sedang}|C1) = 3/7 = 0.25$
 - c) $P(\text{Unlimited Ya}|C1) = 2/7 = 0.28$
 - d) $P(\text{Jangka Aktif 1 Bulan}|C1) = 5/7 = 0.71$
 - e) $P(\text{Provider XL}) = 2/7 = 0.28$
2. Hitung peluang seluruh atribut terhadap *Class* 1
 $P(X|C1) = \prod P(X_k|C1) = 0.43 * 0.25 * 0.28 * 0.71 * 0.28 = 0.0059$
3. Hitung peluang *C1*
 $P(C1) = 7/20 = 0.35$
4. Hitung peluang *C1* berdasarkan *case*
 $P(C1|X) = 0.35 * 0.0059 = \mathbf{0.002065}$

b. *Class* 2: Minat Tinggi

1. Hitung peluang setiap atribut *case* terhadap *Class* 2
 - a) $P(\text{Kuota Besar}|C2) = 0$
 - b) $P(\text{Harga Sedang}|C2) = 0$
 - c) $P(\text{Unlimited Ya}|C2) = 0$
 - d) $P(\text{Jangka Aktif 1 Bulan}|C2) = 2/2 = 1$
 - e) $P(\text{Provider XL}) = 2/2 = 1$
2. Hitung peluang seluruh atribut terhadap *Class* 2
 $P(X|C2) = \prod P(X_k|C2) = 0 * 0 * 0 * 1 * 1 = 0$
 Hitung peluang *C2*
 $P(C2) = 2/20 = 0.1$
 Hitung peluang *C2* berdasarkan *case*
 $P(C2|X) = 0.1 * 0 = 0$
3. Hitung peluang setiap atribut *case* terhadap *Class* 3
 - a) $P(\text{Kuota Besar}|C3) = 3/11 = 0.27$
 - b) $P(\text{Harga Sedang}|C3) = 5/11 = 0.45$
 - c) $P(\text{Unlimited Ya}|C3) = 1/11 = 0.09$

$$P(\text{Jangka Aktif 1 Bulan}|C3) = 10/11 = 0.90$$

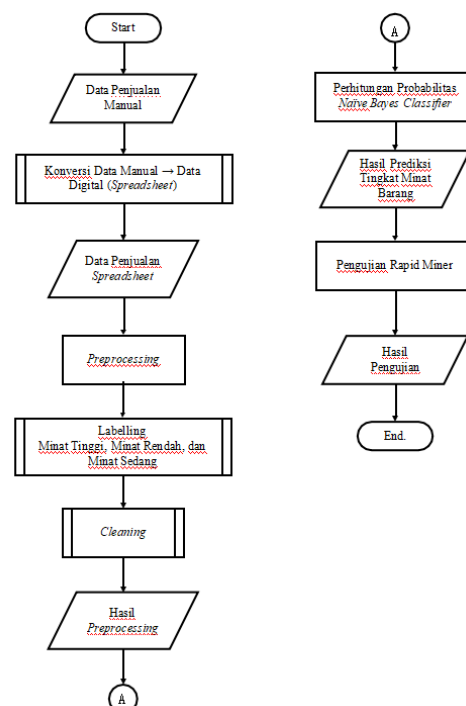
$$P(\text{Provider XL}) = 3/11 = 0.27$$

4. Hitung peluang seluruh atribut terhadap *Class* 3
 $P(X|C2) = \prod P(X_k|C2) = 0.27 * 0.45 * 0.09 * 0.90 * 0.27 = 0.0026$
 Hitung peluang *C3*
 $P(C2) = 11/20 = 0.55$
5. Hitung peluang *C3* berdasarkan *case*
 $P(C2|X) = 0.1 * 0 = 0.00143$

Hasil *Naïve Bayes Classifier* berdasarkan perhitungan tersebut menunjukkan bahwa *Class* 1 (Minat Rendah) memiliki peluang yang lebih besar daripada *Class* 2 (Minat Tinggi) dan *Class* 3 (Minat Sedang) dengan hasil 0.002065.

3.7. Pengujian Akurasi Algoritma

Setelah itu, hasil klasifikasi tersebut akan diuji tingkat akurasi dengan membandingkan hasil klasifikasi menggunakan *Rapid Miner*. Secara keseluruhan, maka alur-alur pada penelitian ini dapat dilihat pada Gambar 2 berikut



Gambar 2. Alur Penelitian Analisis Algoritma *Naive Bayes Classifier* (NBC) pada Klasifikasi Tingkat Minat Barang di Toko Violet Cell

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data penjualan yang berhasil diperoleh adalah data penjualan untuk periode bulan Mei 2022 hingga Juni 2022 sejumlah 144 transaksi. Data tersebut akan dibagi menjadi dua bagian yaitu data latih dan data uji

dengan komposisi 20%:80%. Data yang akan diuji yaitu sebesar 20% dari total data keseluruhan yaitu sejumlah 29 data dari 144 data. Berikut adalah beberapa contoh data transaksi yang telah dikumpulkan oleh penulis.

Tabel 4. Contoh data penjualan

Tgl.	Nama	Kuota (gb)	Harga	Terjual
01/05/2022	Voucher XL S 9GB/BULAN	9	25000	7
02/05/2022	Kartu Perdana Indosat Freedom U 7GB+20GB Apps / 30 Hari	27	85000	7
02/05/2022	Voucher XL XTRA ON 4GB/BULAN	4	20000	15
02/05/2022	Kartu Perdana Indosat Freedom U 7GB+20GB Apps / 30 Hari	27	85000	7
02/05/2022	Voucher XL XTRA ON 4GB/BULAN	4	20000	15

4.2. Preprocessing Data

Data yang telah dikumpulkan tersebut kemudian dilakukan *preprocessing* berupa *labelling* dan *cleaning*. *Preprocessing* dilakukan untuk menyiapkan data agar dapat diolah pada proses penerapan algoritma *Naïve Bayes Classifier*. Selain itu, data transaksi tersebut juga akan digunakan untuk proses pengujian menggunakan *Rapid Miner*.

a. Labelling

Pada proses ini, data akan dikategorikan berdasarkan besaran kuota, kategori harga, dan tingkat penjualannya sesuai dengan kategori yang telah ditentukan. Tabel 5 merupakan tolok ukur *labelling* sehingga dapat dilakukan *labelling* pada setiap data penjualan.

Tabel 5. Konversi Data Penjualan

Kategori	Rendah	Sedang	Tinggi
Kuota	Kuota ≤ 5 Gb	5 Gb < Kuota ≤ 10 Gb	Kuota > 10 Gb
Harga	Harga $\leq$ Rp 25.000	25.000 < Harga $\leq$ Rp 25.001	Harga > Rp 50.000
Jumlah Terjual	Jumlah ≤ 5 buah	5 < Jumlah ≤ 10 buah	Jumlah > 10 buah

Berikut adalah sampel hasil labelling data penjualan berdasarkan pengkategorian tersebut.

Tabel 6. Sampel Hasil Labelling Data Penjualan

Nama Item	Kuota	Kategori Kuota	Harga	Kategori Harga	Terjual	Minat
Voucher XL S 9GB/BULAN	9	Besar	25000	Murah	7	Sedang
Kartu Perdana Indosat Freedom U 7GB+20GB Apps / 30 Hari	27	Kecil	85000	Mahal	7	Sedang
Voucher XL XTRA ON 4GB/BULAN	4	Besar	20000	Murah	15	Tinggi

b. Cleaning

Setelah data diberi label, maka kolom-kolom data yang tidak digunakan untuk klasifikasi dihapus. Tabel 7 adalah sampel hasil proses *cleaning* data setelah *labelling*.

Tabel 7. Sampel Hasil Cleaning Data Penjualan

Provider	Kategori Kuota	Kategori Harga	Minat
XL	Besar	Murah	Sedang
INDOSAT	Kecil	Mahal	Sedang
XL	Besar	Murah	Tinggi
INDOSAT	Kecil	Mahal	Sedang
XL	Besar	Murah	Tinggi

4.3. Penerapan Algoritma Naive Bayes

Tahap selanjutnya yaitu penerapan algoritma *Naïve Bayes Classifier* (NBC) untuk klasifikasi tingkat minat paket kuota internet pada Toko Violet Cell berdasarkan *dataset* yang telah diolah pada *preprocessing*.

Gambar 3. Tampilan Halaman Aplikasi Klasifikasi Tingkat Minat

Gambar 3 tersebut merupakan halaman awal yang muncul saat pengguna membuka atau mengakses berit tersebut. Terdapat empat atribut yang dapat diisi oleh user sesuai dengan kebutuhan klasifikasi, yaitu Kategori Harga, Kuota, *Unlimited*, dan *Provider*.

Percobaan klasifikasi tingkat minat barang dilakukan dengan melakukan uji coba memilih *case* secara acak untuk mengetahui kategori *class* yang terbaik. Datayang akan diuji yaitu sebesar 20% dari total data keseluruhan yaitu sejumlah 29 data dari 144 data. Berikut adalah contoh penerapan klasifikasi

tingkat minat barang menggunakan aplikasi berbasis web.

Gambar 4. Penerapan Algoritma pada Aplikasi Berbasis Web

Secara keseluruhan, berikut ini adalah tabel hasil pengujian klasifikasi tingkat minat barang di Toko Violet Cell menggunakan *Naïve Bayes Classifier*.

Berdasarkan tabel 29 kali percobaan didapatkan 5 prediksi yang memiliki kesalahan prediksi, sedangkan 24 data lainnya diprediksi dengan benar. Ketidaksesuaian hasil prediksi antara minat berdasarkan penjualan dan minat berdasarkan klasifikasi *Naïve Bayes Classifier* disebabkan oleh atribut yang digunakan untuk melakukan klasifikasi.

Pada klasifikasi *Naïve Bayes Classifier* digunakan 4 atribut yaitu *provider*, besar kuota, kategori harga, dan status limit kuota tersebut sehingga hasilnya berbeda dengan klasifikasi manual yang hanya mengacu pada jumlah penjualan. Selain itu, kesalahan klasifikasi tersebut juga dapat disebabkan oleh kurangnya data yang dijadikan sebagai data latih.

Tabel 8. Hasil Klasifikasi Tingkat Minat Barang dengan NBC

No	Provider	Kategori Kuota	Kategori Harga	Unlimited	Minat Berdasarkan Penjualan	Hasil Klasifikasi NBC
1	Indosat	Kecil	Mahal	Tidak	Rendah	Rendah
2	Tri Aon	Kecil	Mahal	Tidak	Rendah	Rendah
3	Tri Aon	Kecil	Mahal	Tidak	Rendah	Rendah
4	Tri Aon	Kecil	Mahal	Tidak	Rendah	Rendah
5	Tri Aon	Kecil	Mahal	Ya	Rendah	Rendah
6	Tri	Kecil	Murah	Ya	Rendah	Rendah
7	Tri	Kecil	Murah	Tidak	Rendah	Rendah
8	Telkomsel	Kecil	Mahal	Tidak	Rendah	Rendah
9	Indosat	Kecil	Murah	Tidak	Rendah	Sedang
10	Telkomsel	Kecil	Mahal	Ya	Rendah	Rendah
11	Indosat	Kecil	Murah	Ya	Rendah	Sedang
12	Xl	Kecil	Mahal	Ya	Rendah	Rendah
13	Indosat	Kecil	Sedang	Ya	Rendah	Rendah
14	Xl	Kecil	Mahal	Tidak	Rendah	Rendah
15	Xl	Kecil	Mahal	Tidak	Rendah	Rendah
16	Tri Aon	Kecil	Murah	Tidak	Rendah	Sedang
17	Smartfren	Kecil	Mahal	Tidak	Rendah	Rendah
18	Tri Aon	Kecil	Sedang	Tidak	Rendah	Rendah
19	Smartfren	Kecil	Mahal	Tidak	Rendah	Rendah
20	Tri Aon	Kecil	Mahal	Tidak	Rendah	Rendah
21	Tri Aon	Kecil	Sedang	Tidak	Rendah	Rendah
22	Smartfren	Kecil	Murah	Tidak	Rendah	Sedang
23	Smartfren	Kecil	Sedang	Tidak	Rendah	Rendah
24	Smartfren	Kecil	Mahal	Tidak	Rendah	Rendah
25	Indosat	Kecil	Mahal	Tidak	Rendah	Rendah
26	Indosat	Kecil	Sedang	Tidak	Rendah	Rendah
27	Smartfren	Kecil	Sedang	Tidak	Rendah	Rendah
28	Smartfren	Kecil	Murah	Tidak	Tinggi	Sedang
29	Xl	Kecil	Murah	Tidak	Tinggi	Tinggi

4.4. Pengujian Akurasi

Pengujian akurasi klasifikasi ini menggunakan aplikasi *Rapid Miner* dengan cara membandingkan hasil klasifikasi dari aplikasi buatan penulis dengan hasil klasifikasi menggunakan *Rapid Miner*. Pada pengujian menggunakan *Rapid Miner*, data akan

dibagi menjadi dua bagian, bagian pertama sebesar 80% untuk pelatihan model dan bagian kedua sebesar 20% untuk pengujian.

Tabel 9. Hasil Klasifikasi Tingkat Minat Barang dengan NBC Menggunakan *Rapid Miner*

Provider	Kategori Kuota	Kategori Harga	Unlimited	Minat	Confidence (Sedang)	Confidence (Tinggi)	Confidence (Rendah)	Prediction (Minat)
INDOSAT	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Mahal	Ya	Rendah	0,1	0,0	0,9	Rendah
Tri	Kecil	Murah	Ya	Rendah	0,0	0,0	1,0	Rendah
Tri	Kecil	Murah	Tidak	Rendah	0,0	0,0	1,0	Rendah
Telkomsel	Kecil	Mahal	Tidak	Rendah	0,0	0,0	1,0	Rendah
INDOSAT	Kecil	Murah	Tidak	Rendah	0,7	0,0	0,3	Sedang
Telkomsel	Kecil	Mahal	Ya	Rendah	0,0	0,0	1,0	Rendah
INDOSAT	Kecil	Murah	Ya	Rendah	0,8	0,0	0,2	Sedang
XL	Kecil	Mahal	Ya	Rendah	0,2	0,0	0,8	Rendah
INDOSAT	Kecil	Sedang	Ya	Rendah	0,3	0,0	0,7	Rendah
XL	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
XL	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Murah	Tidak	Rendah	0,6	0,0	0,4	Sedang
Smartfren	Kecil	Mahal	Tidak	Rendah	0,0	0,0	1,0	Rendah
Tri AON	Kecil	Sedang	Tidak	Rendah	0,1	0,0	0,9	Rendah
Smartfren	Kecil	Mahal	Tidak	Rendah	0,0	0,0	1,0	Rendah
Tri AON	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
Tri AON	Kecil	Sedang	Tidak	Rendah	0,1	0,0	0,9	Rendah
Smartfren	Kecil	Murah	Tidak	Rendah	0,4	0,3	0,3	Sedang
Smartfren	Kecil	Sedang	Tidak	Rendah	0,1	0,0	0,8	Rendah
Smartfren	Kecil	Mahal	Tidak	Rendah	0,0	0,0	1,0	Rendah
INDOSAT	Kecil	Mahal	Tidak	Rendah	0,1	0,0	0,9	Rendah
INDOSAT	Kecil	Sedang	Tidak	Rendah	0,3	0,0	0,7	Rendah
Smartfren	Kecil	Sedang	Tidak	Rendah	0,1	0,0	0,8	Rendah
Smartfren	Kecil	Murah	Tidak	Tinggi	0,4	0,3	0,3	Sedang
XL	Kecil	Murah	Tidak	Tinggi	0,4	0,5	0,1	Tinggi

Berdasarkan hasil pengujian yang didapat oleh penulis, didapatkan hasil klasifikasi yang ditunjukkan aplikasi *Rapid Miner* dan aplikasi web oleh penulis memiliki hasil yang sama. Terdapat 5 data yang memiliki kesalahan prediksi sama dengan hasil klasifikasi *Naïve Bayes Classifier* menggunakan

aplikasi berbasis web. Secara keseluruhan maka hasil klasifikasi tingkat minat barang menggunakan algoritma *Naïve Bayes Classifier* adalah 23 barang memiliki kategori “Minat Rendah”, 5 barang dikategorikan “Minat Sedang”, dan 1 barang memiliki kategori “Minat Tinggi”.

accuracy: 82.76%				
	true Rendah	true Tinggi	true Sedang	class precision
pred. Rendah	23	0	0	100.00%
pred. Tinggi	0	1	0	100.00%
pred. Sedang	4	1	0	0.00%
class recall	85.19%	50.00%	0.00%	

Gambar 5. Hasil *Confusion Matrix*

Berdasarkan Gambar 5, hasil klasifikasi menggunakan algoritma *Naïve Bayes Classifier* memperoleh tingkat *accuracy* sebesar 82,76%. Skor akurasi tersebut diperoleh dengan menghitung jumlah data yang diklasifikasi secara benar dibagi dengan keseluruhan data. Pada gambar 4.3 dapat dilihat bahwa tingkat *recall* untuk klasifikasi Rendah sebesar 85,19%, Tinggi sebesar 50%, dan Sedang 0%. Selain itu, klasifikasi algoritma *Naïve Bayes Classifier* memperoleh *precision* sebesar 100% untuk prediksi *class* Rendah, 100% untuk prediksi *class* Tinggi dan 0% untuk prediksi *class* Sedang.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, diperoleh hasil sebagai berikut, klasifikasi tingkat minat barang yang ditunjukkan aplikasi web dan aplikasi *rapid miner* memiliki hasil yang sama. Terdapat 5 data yang memiliki kesalahan prediksi sama dengan hasil klasifikasi *naïve bayes classifier* menggunakan aplikasi berbasis web. Secara keseluruhan maka hasil klasifikasi tingkat minat barang menggunakan algoritma *naïve bayes classifier* adalah 23 barang memiliki kategori “minat rendah”, 5 barang dikategorikan “minat sedang”, dan 1 barang memiliki kategori “minat tinggi”. Pengujian akurasi klasifikasi tingkat barang menggunakan aplikasi *rapid*

miner dengan cara membandingkan hasil klasifikasi dari aplikasi buatan penulis dengan hasil klasifikasi menggunakan *rapid miner*. Pada pengujian menggunakan *rapid miner* data akan dibagi menjadi dua bagian, bagian pertama sebesar 80% untuk pelatihan model dan bagian kedua sebesar 20% untuk pengujian. Berdasarkan perhitungan *confusion matrix* hasil klasifikasi menggunakan *algoritma naïve bayes classifier* memperoleh tingkat akurasi sebesar 82,76%, *recall* untuk klasifikasi rendah sebesar 85,19%, tinggi sebesar 50%, dan sedang 0%. Selain itu, klasifikasi *algoritma naïve bayes classifier* memperoleh *precision* sebesar 100% sedangkan untuk prediksi *class* rendah, 100% untuk prediksi *class* tinggi dan 0% untuk prediksi *class* sedang. Peneliti berharap pada peneliti selanjutnya dapat menggunakan sampel transaksi yang lebih banyak agar tingkat akurasi meningkat serta aplikasi dapat ditambahkan beberapa fitur pendukung yang dibutuhkan oleh pengguna, seperti integrasi data penjualan dan data transaksi.

DAFTAR PUSTAKA

- [1] Sikumbang, E.D.. 2018. Penerapan *Data Mining* Penjualan Sepatu Menggunakan Metode Algoritma Apriori. Jurnal Teknik Komputer Vol 4, No. 1.
- [2] Nofriansyah, D., Erwansyah, K., Ramadhan, M.. 2016. Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi). Jurnal Saintikom Vol. 15, Hal. 81-92.
- [3] Ota, L.. 2016. Probabilitas Bersyarat, Independensi dan Teorema Bayes Dalam Menentukan Peluang Terjadinya Suatu Peristiwa. Jurnal Manajemen Pendidikan Islam Vol. 4., Hal. 68-78.
- [4] Retnoningsih, E., Pramudita, R.. 2020. Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. BINA INSANI ICT JOURNAL Vol.7, No.2, Desember 2020, Hal. 156-165.
- [5] Rerung, R.R.. 2018. Penerapan *Data Mining* dengan Memanfaatkan Metode *Association Rule* untuk Promosi Produk. Jurnal Teknologi Rekayasa, Vol. 3, No. 1, Juni 2018, Hal. 89-98
- [6] Prastetya, Ratih. 2020. Penerapan Teknik Data Mining Dengan Algoritma *Classification Tree* Untuk Prediksi Hujan. Jurnal Widya Climago, Vol.2 No.2, November 2020.
- [7] Widiari, Ni P.A., Suarjaya, I. M.A.D., Githa, D.P.. 2020. Teknik *Data Cleaning* Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali. Jurnal Ilmiah Merpati Vol. 8, No. 2 August 2020
- [8] Fitriani, Arif S.. 2019. Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur. JTAM (Jurnal Teori dan Aplikasi Matematika) Vol. 3, No. 2, Oktober 2019, Hal. 98-104.
- [9] Irawan, Yuda. 2021. Penerapan Algoritma *Decision Tree* C4.5 Untuk Prediksi Kelayakan Calon Pendonor Darah Dengan Klasifikasi Data Mining. JTAM : Jurnal Teknologi Informasi dan Multimedial-ISSN : 2715-2529 Vol. 2, No. 4, Pebruari2021, hlm. 181-189.
- [10] Susana, Heliyanti, Suarna, Nana., Fathurrohman, Kaslani. 2022. Penerapan Model Klasifikasi Metode *Naive Bayes* Terhadap Penggunaan Akses Internet. JURSISTEKNI (Jurnal Sistem Informasi dan Teknologi Informasi) Vol 4, No.1, January 2022: Hal 1– 8 ISSN.
- [11] Chamdani, Muhammad N. 2022. Probabilitas & Statistika. Jurusan Teknik Elektro Universitas Nusa Putra, Oktober , 2022.
- [12] Siregar, N.C., Siregar, Riki R.A., Sudirman, M. Y.D.. 2020. Implementasi Metode *Naive Bayes Classifier (NBC)* Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ). Jurnal Teknologia Aliansi Perguruan Tinggi (APERTI) BUMN Vol. 3, No. 1, Agustus 2020, ISSN 2654-568.