

# Klasifikasi Pengguna Shopee Berdasarkan Promosi Menggunakan Naïve Bayes

Tania Fatiah Rahmadanti<sup>1</sup>, Mohamad Jajuli<sup>2</sup>, Intan Purnamasari<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

E-mail: <sup>1</sup>[tania.fatihah17202@student.unsika.ac.id](mailto:tania.fatihah17202@student.unsika.ac.id), <sup>2</sup>[mohamad.jajuli@unsika.ac.id](mailto:mohamad.jajuli@unsika.ac.id),

<sup>3</sup>[intan.purnamasari@staff.unsika.ac.id](mailto:intan.purnamasari@staff.unsika.ac.id)

**Abstrak** – Belanja online merupakan transaksi jual beli barang atau jasa melalui media perantara yaitu jejaring sosial. Faktor yang mempengaruhi terjadinya perubahan pola konsumsi dan cara belanja masyarakat yang awalnya konvensional beralih melalui layanan E-Commerce yaitu, meningkatnya minat masyarakat dalam berbelanja online yang disebabkan adanya wabah virus COVID-19, dan sepanjang tahun 2019 pengguna E-Commerce yang melakukan transaksi online mencapai 168.3 juta orang. Berdasarkan data laporan iprice tahun 2020 Shopee merupakan E-Commerce yang paling banyak dikunjungi dengan total pengunjung 129.320.800. Shopee merupakan pihak ketiga yang menyediakan tempat berjualan dan fasilitas pembayaran maka dari itu Shopee tidak bertanggung jawab atas pemasaran produk yang dijual. Untuk menarik minat konsumen penjual memerlukan promosi yang menarik. Oleh karena itu diperlukan penelitian untuk mengklasifikasikan pengguna E-Commerce. Tujuan penelitian untuk mengklasifikasikan pengguna E-Commerce ini berdasarkan promosi yang dipakai menggunakan algoritma Naïve Bayes dengan metodologi Knowledge Discovery in Database (KDD). Dilakukan sembilan skenario pengujian dengan cross validation yang menunjukkan bahwa performa terbaik yaitu skenario pengujian dengan menggunakan 3 fold yang menghasilkan performa dengan nilai accuracy sebesar 88,73%, dan dengan nilai kappa yang diperoleh sebesar 0,451 yang termasuk kedalam kategori cukup (moderate). Berdasarkan hasil tersebut maka, model yang dihasilkan oleh algoritma Naïve Bayes ini cukup konsisten.

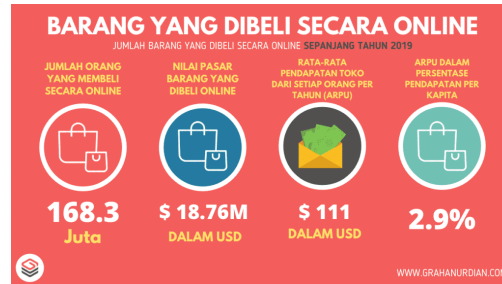
**Kata Kunci** — E-Commerce, Klasifikasi, Naïve Bayes, Promosi

**Abstract** – Online shopping is a transaction of buying and selling goods or services through intermediary media, namely social networks. There has been a change in consumption patterns and the way people spend their money, which was originally conventional to switch to E-Commerce services due to several factors, namely the increasing public interest in online shopping due to the COVID-19 virus outbreak, and throughout 2019 E-Commerce users who made transactions reached 168.3 million people. . Based on iprice report data in 2020, Shopee is the most visited E-Commerce with a total of 129,320,800 visitors. Shopee is only a third party that provides a place to sell and payment facilities, therefore Shopee is not responsible for marketing the products sold. To attract consumers, sellers need attractive promotions. Therefore, research is needed to classify E-Commerce users. The purpose of this research is to classify E-Commerce users based on the promotion used using the Naïve Bayes algorithm with the Knowledge Discovery in Database (KDD) methodology. Nine test scenarios were carried out with cross validation which showed that the best performance was a test scenario using 3 folds which resulted in performance with an accuracy value of 88.73%, and with a kappa value of 0.451 which was included in the moderate category. Based on these results, the model generated by the Naïve Bayes algorithm is quite consistent.

**Keywords** — Classification, E-Commerce, Naïve Bayes, Promotion

## 1. PENDAHULUAN

Generasi muda atau milenial tentunya tidak asing dengan namanya belanja *online*. Belanja *online* merupakan transaksi jual beli barang atau jasa melalui media perantara yaitu jejaring sosial [1]. Banyak keuntungan dari belanja *online*, diantaranya tidak perlu mendatangi toko secara langsung, banyak promosi yang ditawarkan, pilihan metode pembayaran yang beragam, dan tidak perlu mengantri. Faktor lain yang mengakibatkan meningkatnya minat masyarakat dalam berbelanja *online* yaitu karena adanya wabah virus *COVID-19*, sehingga pemerintah mengeluarkan kebijakan untuk membatasi aktifitas keluar rumah [2]. Dari faktor tersebutlah terjadinya perubahan pola konsumsi dan cara belanja masyarakat yang awalnya konvensional beralih melalui layanan *E-Commerce*.



Gambar 1. Pembelian Barang Secara Online (Sumber: Graha Nurdian, 2021)

Faktor lainnya juga dapat dilihat pada gambar 1 yaitu sepanjang tahun 2019 pengguna *E-Commerce* yang melakukan transaksi sebanyak 168.3 juta orang. Nilai pasaran barang yang dibeli secara *online* pun mencapai \$18.76M. Bahkan rata-rata toko *online* memiliki pendapatan sebanyak \$111 dari setiap orang selama satu tahun [3].



Gambar 2. *E-Commerce* yang Sering Dikunjungi Tahun 2020

Terdapat beberapa *E-Commerce* yang dapat digunakan diantaranya yaitu Bukalapak, Tokopedia, Lazada, Shopee, Blibli, Tiket.com, Kaskus, O LX, dan lain-lain. Berdasarkan data laporan *iprice* Shopee menempati peringkat pertama *E-Commerce* yang paling sering di kunjungi dengan total 129.320.800 pengunjung [4]. Banyaknya penjual yang menggunakan Shopee untuk memasarkan produknya dan memberikan banyak promosi sehingga banyak masyarakat yang berbelanja *online* melalui Shopee. Shopee juga dapat dengan mudah diterima oleh masyarakat karena dalam sifat penggunaannya Shopee termasuk kedalam kelompok *Customer to Customer* (C2C). *Customer to Customer* (C2C) merupakan bisnis yang dimana konsumen menjual langsung kepada konsumen [5]. Shopee juga menyediakan berbagai macam produk yang dibutuhkan oleh masyarakat dengan pilihan harga terjangkau dan beragam. Shopee merupakan *E-Commerce* atau pihak ketiga yang menyediakan tempat berjualan dan fasilitas pembayaran, tetapi Shopee tidak bertanggung jawab atas pemasaran produk yang dijual. Oleh karena itu para penjual harus bersaing untuk menarik minat para konsumen. Untuk menarik minat konsumen penjual memerlukan promosi yang menarik. Promosi yang dapat ditawarkan di Shopee dapat berupa voucher diskon, *cashback*, gratis ongkos kirim, dan *flash sale*.

Data *mining* merupakan proses pemanfaatan data dengan jumlah yang besar yang kemudian diolah untuk mendapatkan informasi penting [6]. Untuk membantu penjual dalam menentukan promosi, maka dilakukan pengimplementasian data *mining* dengan mengklasifikasikan pengguna

berdasarkan promosi yang digunakannya. Data *mining* dengan teknik klasifikasi telah digunakan pada penelitian sebelumnya oleh Ranggadara (2019) dalam klasifikasi penentuan pemberian *reward* yang menghasilkan akurasi sebesar 87% [7]. Romli (2019) menggunakan teknik klasifikasi untuk menentukan tingkat penjualan mobil menggunakan algoritma Naïve Bayes. Pada penelitian tersebut menghasilkan nilai akurasi sebesar 96.87% [8]. Klasifikasi dengan algoritma Naïve Bayes juga dilakukan oleh Putro (2020) untuk mengklasifikasikan pelanggan yang berpotensi atau tidak untuk dasar pertimbangan keputusan yaitu strategi pemasaran [9].

Dari penelitian yang telah dilakukan sebelumnya, algoritma Naïve Bayes memiliki akurasi yang besar pada jumlah data yang sedikit [10]. Sehingga pada penelitian ini juga akan menggunakan metode Naïve Bayes. Perbedaan pada penelitian sebelumnya terletak pada kelas atau label yang digunakan. Penelitian sebelumnya mengklasifikasikan pelanggan kedalam 2 kelas saja, sedangkan penelitian yang dilakukan sekarang mengklasifikasikan pelanggan Shopee berdasarkan 5 kelas promosi yang digunakannya. Evaluasi yang dilakukan pada penelitian sebelumnya dilihat dari nilai akurasi, sedangkan evaluasi yang dilakukan pada penelitian ini menggunakan nilai akurasi dan juga nilai *kappa*.

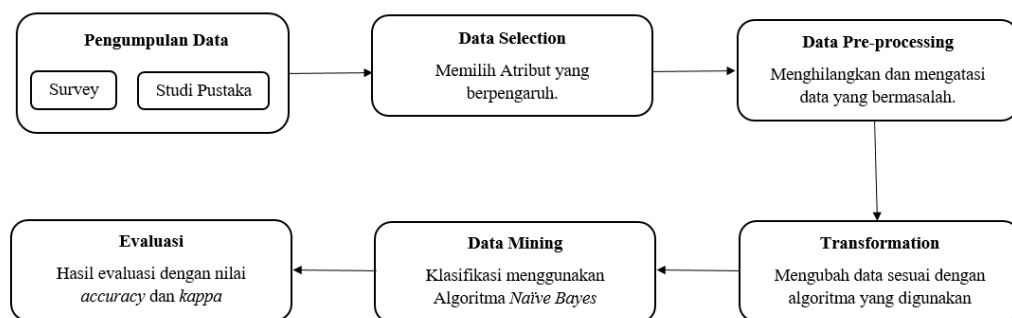
## 2. METODE PENELITIAN

### 2.1. Objek Penelitian

Objek penelitian yang akan diteliti yaitu pengguna Shopee. Pengelompokan ini dilakukan kepada pengguna Shopee yang telah melakukan transaksi.

### 2.2. Metodologi Penelitian

Penelitian ini menggunakan metodologi *Knowledge Discovery in Database* (KDD) yang dimana memiliki alur penelitian sebagai berikut.



Gambar 3. Alur Penelitian

#### 2.2.1. Pengumpulan Data

Pengumpulan data dilakukan dengan survei dan studi pustaka. Studi pustaka dilakukan untuk mendapatkan informasi relevan yang dapat mendukung objek penelitian. Informasi yang didapatkan dapat diperoleh dari karya ilmiah, buku, jurnal, tesis, disertasi, internet, dan sumber lainnya yang berkaitan dengan klasifikasi data *mining* dan algoritma Naïve Bayes. Dengan studi pustaka penelitian lebih ditunjang dari teori dan bukti nyata.

Teknik pengumpulan data pada penelitian ini adalah survei. Survei ialah penelitian yang menggunakan sampel dari satu populasi dan memakai alat pendataan berupa kuesioner [11]. Survei dilakukan dengan membagikan angket/kuesioner melalui *google form* untuk responden yang telah melakukan transaksi *E-Commerce* pada Shopee. Jenis kuesioner yang akan digunakan yaitu kuesioner campuran dimana jawaban yang dibutuhkan telah disediakan sebelumnya dengan tipe pilihan ganda dan juga lisan.

### 2.2.2. Data Selection

*Data selection* dilakukan untuk memilih atribut yang berpengaruh untuk proses klasifikasi pengguna Shopee. Untuk melakukan proses klasifikasi diperlukan atribut-atribut yang berpengaruh terhadap klasifikasi pengguna Shopee. Atribut yang berpengaruh dalam pengklasifikasian pengguna Shopee yaitu, usia, jenis kelamin, mengetahui shopee, memiliki akun shopee, sebagai apa di shopee, rata-rata pembelian, menggunakan promosi, promosi yang digunakan, produk, dan pembayaran. Klasifikasi pengguna Shopee dikategorikan kedalam kelas diskon, gratis ongkir, *flash sale*, *cashback*, dan kategori tidak ada jika tidak menggunakan promosi.

### 2.2.3. Data Pre-Processing

Tahapan ini dilakukan untuk mengubah data mentah menjadi data yang berkualitas. Data *pre-processing* dilakukan untuk menghilangkan atau mengatasi data yang bermasalah seperti, *missing value*, data eror, data yang tidak valid, dan data yang tidak konsisten. Untuk mengatasi data yang bermasalah dapat dengan mengganti nilai dengan rata-rata pada kelas yang sama (*replace missing value*) atau menghapus data yang terdapat *missing value* (*drop missing value*) [12].

### 2.2.4. Transformation

*Transformation* dilakukan untuk mengubah data sesuai dengan bentuk yang dibutuhkan. Proses *transformasi* pada penelitian ini dilakukan dengan teknik *discretization*. Teknik tersebut mengubah atribut numerik menjadi atribut kategorikal [13]. Perubahan ini dilakukan untuk mengkategorikan atribut numerik menjadi beberapa tingkatan kategori. *Transformation* pada penelitian ini dilakukan pada atribut usia dan rata-rata pembelian dengan teknik *Discretization by Binning*.

### 2.2.5. Data Mining

*Data mining* dilakukan untuk mencari pola atau informasi yang unik pada sebuah data yang sudah dipilih dengan menggunakan teknik atau cara tertentu [14]. Pada tahapan *data mining* ini dilakukan klasifikasi pengguna Shopee dengan metode Naïve Bayes. Alur kerja dari metode Naïve Bayes, yaitu pertama membaca data training yang sudah siap, lalu menghitung jumlah kelas atau label, menghitung jumlah kasus per kelas, selanjutnya mengkalikan semua variabel kelas dan bandingkan hasil per kelas. Hasil penentuan kelas dilihat dari probabilitas yang paling tinggi.

### 2.2.6. Evaluasi

Tahapan evaluasi dilakukan untuk menilai sejauh mana model yang telah dibuat sehingga dapat memenuhi tujuan klasifikasi dan mengevaluasi hasil penelitian dari nilai *accuracy* dan nilai *kappa*.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Metode Naïve Bayes

Naïve Bayes ialah pengklasifikasian statistik yang dipakai untuk memprediksi probabilitas keanggotaan sebuah *class* [15]. Pengklasifikasian akan dilaksanakan perhitungan nilai probabilitas dari setiap label kelas yang ada terhadap masukan yang diberikan. Label kelas diambil dari nilai probabilitas yang sangat besar.

Bentuk umum dari teorema *bayes* yaitu [16] :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data merupakan suatu kelas spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posterior probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H  
 P(X) : Probabilitas X

3.1.1. Menghitung Jumlah Kelas

Dari 284 data yang siap digunakan diketahui pada atribut yang menjadi kelas terdapat data dengan kelas “Gratis Ongkir” sebanyak 239 data, “Flash Sale” sebanyak 8 data, “Cashback” sebanyak 12 data, “Diskon” sebanyak 13 data, dan “Tidak Ada” yang menggunakan promosi sebanyak 12 data. Data dari masing-masing kelas dibagi dengan keseluruhan data maka akan mendapatkan probabilitas prior P(X). Perhitungan probabilitas prior untuk kemungkinan kelas gratis ongkir, flash sale, cashback, diskon, dan kelas tidak ada didapatkan sebagai berikut :

$$P(\text{Gratis Ongkir}) = 239/284 = 0.842$$

$$P(\text{Flash Sale}) = 8/284 = 0.028$$

$$P(\text{Cashback}) = 12/284 = 0.042$$

$$P(\text{Diskon}) = 13/284 = 0.046$$

$$P(\text{Tidak Ada}) = 12/284 = 0.042$$

3.1.2. Menghitung Jumlah Kasus Dari Setiap Kelas

Setelah menghitung probabilitas dari setiap kelas, selanjutnya menghitung probabilitas posterior atau jumlah kasus maupun kejadian dari setiap kelas. Untuk mencari nilai probabilitas yaitu jumlah atribut dengan kelas “Gratis Ongkir”, “Flash Sale”, “Cashback”, “Diskon”, dan kelas “Tidak Ada” kemudian dibagi dengan jumlah kelas yang ada.

Tabel 1. Jumlah Pengguna Shopee Seluruh Atribut Dari Setiap Kelas

Atribut	Nilai Atribut	Jumlah Kejadian				
		Gratis Ongkir	Flash Sale	Cash-back	Diskon	Tidak Ada
Usia	≤20	113	2	5	12	4
	>20	126	6	7	1	8
Jenis Kelamin	Perempuan	211	4	8	12	7
	Laki-Laki	28	4	4	1	5
Mengetahui Shopee	Ya	239	8	12	13	12
	Tidak	0	0	0	0	0
Memiliki Akun Shopee	Ya	239	7	12	13	12
	Tidak	0	1	0	0	0
Sebagai Apa Di Shopee	Pembeli	221	6	12	11	12
	Penjual	18	2	0	2	0
Rata-Rata Pembelian	Jarang	235	6	12	12	12
	Biasa	2	2	0	1	0
	Sering	2	0	0	0	0
Menggunakan Promosi Produk	Ya	239	8	12	13	0
	Tidak	0	0	0	0	12
	Elektronnik	10	4	0	0	1
	Fashion	85	1	4	4	5
	Makanan/Minuman	6	0	0	0	0
	Perawatan&Kecantikan	95	3	5	6	3
	Kebutuhan Rumah Tangga	13	0	1	1	0
	Kesehatan Lainnya	2	0	0	1	0
Pembayaran	ShopeePay/ ShopeePayLater	171	7	9	12	5
	Alfamart / Indomaret	25	1	0	1	2
	Transfer Bank	22	0	2	0	4
	COD	19	0	1	0	1
	Kartu Kredit/ Debit Online	2	0	0	0	0

Pada tabel 1 menunjukkan jumlah kejadian pada semua atribut dari setiap kelasnya. Dari total data pada masing-masing atribut didapatkan jumlah kelas “Gratis Ongkir” sebanyak 239 data, “Flash Sale” sebanyak 8 data, “Cashback” sebanyak 12 data, “Diskon” sebanyak 13 data, dan kelas “Tidak Ada” sebanyak 12 data. Untuk mencari nilai probabilitas seluruh atribut dari setiap kelas yang ada yaitu jumlah atribut pada masing-masing kategori dibagi dengan jumlah kelas yang ada. Untuk mencari probabilitas pada atribut usia  $\leq 20$  yang termasuk kedalam kelas gratis ongkir yaitu sebagai berikut,

$$P(\text{Usia} = \text{Usia} \leq 20 | \text{Gratis Ongkir}) = 113/239 = 0,473$$

Sehingga peluang pengguna Shopee yang memiliki kategori usia  $\leq 20$  yang menggunakan promosi gratis ongkir sebesar 0,473. Adapun nilai probabilitas seluruh atribut dari setiap kelasnya dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Probabilitas Seluruh Atribut Dari Setiap Kelas

Atribut	Nilai Atribut	Probabilitas P(X H)				
		Gratis Ongkir	Flash Sale	Cash-back	Diskon	Tidak Ada
Usia	$\leq 20$	0,473	0,25	0,417	0,923	0,333
	$> 20$	0,527	0,75	0,583	0,077	0,667
Jenis Kelamin	Perempuan	0,883	0,5	0,667	0,923	0,583
	Laki-Laki	0,117	0,5	0,333	0,077	0,417
Mengetahui Shopee	Ya	1	1	1	1	1
	Tidak	0	0	0	0	0
Memiliki Akun Shopee	Ya	1	0,875	1	1	1
	Tidak	0	0,125	0	0	0
Sebagai Apa Di Shopee	Pembeli	0,925	0,75	1	0,846	1
	Penjual	0,075	0,25	0	0,154	0
Rata-Rata Pembelian	Jarang	0,983	0,75	1	0,923	1
	Biasa	0,008	0,25	0	0,077	0
	Sering	0,008	0	0	0	0
Menggunakan Promosi	Ya	1	1	1	1	0
	Tidak	0	0	0	0	1
Produk	Elektronik	0,042	0,5	0	0	0,083
	Fashion	0,356	0,125	0,333	0,308	0,417
	Makanan/Minuman	0,025	0	0	0	0
	Perawatan&Kecantikan	0,397	0,375	0,417	0,462	0,25
	Kebutuhan Rumah Tangga	0,054	0	0,083	0,077	0
	Kesehatan	0,008	0	0	0,077	0
	Lainnya	0,117	0	0,167	0,077	0,25
Pembayaran	ShopeePay/ ShopeePayLater	0,715	0,875	0,75	0,923	0,417
	Alfamart / Indomaret	0,105	0,125	0	0,077	0,167
	Transfer Bank	0,092	0	0,167	0	0,333
	COD	0,079	0	0,083	0	0,083
	Kartu Kredit/ Debit Online	0,008	0	0	0	0

### 3.1.3. Mengkalikan Semua Variabel Kelas

Perhitungan nilai probabilitas kelas dan nilai probabilitas kasus dari setiap kelas yang telah dilakukan selanjutnya akan digunakan sebagai model acuan untuk menentukan *data testing*.

Usia	Jenis Kelamin	Mengetahui Shopee	Memiliki Akun Shopee	Sebagai Apa Dalam Shopee	Rata-Rata Pembelian	Menggunakan Promosi	Promosi Yang Digunakan	Produk	Pembayaran
$\leq 20$	Perempuan	Ya	Ya	Pembeli	Jarang	Ya	Gratis Ongkir	Perawatan & Kecantikan	ShopeePay / ShopeePayLater

Gambar 4. Contoh *Data Testing*

Untuk menghitung nilai probabilitas akhir menggunakan persamaan sebagai berikut,

$$P(H|X) = P(X_1|H) \times P(X_2 |H) \times P(X_n|H) \times P(H) \dots\dots\dots(2)$$

1.  $P(\text{Gratis Ongkir} | X_1, X_2, X_3, X_n)$   
 $= P(\text{Gratis Ongkir}) \times P(\text{Usia} = \text{Usia} \leq 20 | \text{Gratis Ongkir}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Gratis Ongkir}) \times P(\text{Mengetahui Shopee} = \text{Ya} | \text{Gratis Ongkir}) \times P(\text{Memiliki Akun Shopee} = \text{Ya} | \text{Gratis Ongkir}) \times P(\text{Sebagai Apa Di Shopee} = \text{Pembeli} | \text{Gratis Ongkir}) \times P(\text{Rata-Rata Pembelian} = \text{Jarang} | \text{Gratis Ongkir}) \times P(\text{Menggunakan Promosi} = \text{Ya} | \text{Gratis Ongkir}) \times P(\text{Produk} = \text{Perawatan \& Kecantikan} | \text{Gratis Ongkir}) \times P(\text{Pembayaran} = \text{ShopeePay/ShopeePayLater} | \text{Gratis Ongkir})$   
 $= 0,842 \times 0,473 \times 0,883 \times 1,000 \times 1,000 \times 0,925 \times 0,983 \times 1,000 \times 0,397 \times 0,715$   
 $= 0,09076653$
2.  $P(\text{Flash Sale} | X_1, X_2, X_3, X_n)$   
 $= P(\text{Flash Sale}) \times P(\text{Usia} = \text{Usia} \leq 20 | \text{Flash Sale}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Flash Sale}) \times P(\text{Mengetahui Shopee} = \text{Ya} | \text{Flash Sale}) \times P(\text{Memiliki Akun Shopee} = \text{Ya} | \text{Flash Sale}) \times P(\text{Sebagai Apa Di Shopee} = \text{Pembeli} | \text{Flash Sale}) \times P(\text{Rata-Rata Pembelian} = \text{Jarang} | \text{Flash Sale}) \times P(\text{Menggunakan Promosi} = \text{Ya} | \text{Flash Sale}) \times P(\text{Produk} = \text{Perawatan \& Kecantikan} | \text{Flash Sale}) \times P(\text{Pembayaran} = \text{ShopeePay/ShopeePayLater} | \text{Flash Sale})$   
 $= 0,028 \times 0,250 \times 0,500 \times 1,000 \times 0,875 \times 0,750 \times 0,750 \times 1,000 \times 0,375 \times 0,875$   
 $= 0,00056525$
3.  $P(\text{Cashback} | X_1, X_2, X_3, X_n)$   
 $= P(\text{Cashback}) \times P(\text{Usia} = \text{Usia} \leq 20 | \text{Cashback}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Cashback}) \times P(\text{Mengetahui Shopee} = \text{Ya} | \text{Cashback}) \times P(\text{Memiliki Akun Shopee} = \text{Ya} | \text{Cashback}) \times P(\text{Sebagai Apa Di Shopee} = \text{Pembeli} | \text{Cashback}) \times P(\text{Rata-Rata Pembelian} = \text{Jarang} | \text{Cashback}) \times P(\text{Menggunakan Promosi} = \text{Ya} | \text{Cashback}) \times P(\text{Produk} = \text{Perawatan \& Kecantikan} | \text{Cashback}) \times P(\text{Pembayaran} = \text{ShopeePay/ShopeePayLater} | \text{Cashback})$   
 $= 0,042 \times 0,417 \times 0,667 \times 1,000 \times 1,000 \times 1,000 \times 1,000 \times 1,000 \times 0,417 \times 0,750$   
 $= 0,00365349$
4.  $P(\text{Diskon} | X_1, X_2, X_3, X_n)$   
 $= P(\text{Diskon}) \times P(\text{Usia} = \text{Usia} \leq 20 | \text{Diskon}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Diskon}) \times P(\text{Mengetahui Shopee} = \text{Ya} | \text{Diskon}) \times P(\text{Memiliki Akun Shopee} = \text{Ya} | \text{Diskon}) \times P(\text{Sebagai Apa Di Shopee} = \text{Pembeli} | \text{Diskon}) \times P(\text{Rata-Rata Pembelian} = \text{Jarang} | \text{Diskon}) \times P(\text{Menggunakan Promosi} = \text{Ya} | \text{Diskon}) \times P(\text{Produk} = \text{Perawatan \& Kecantikan} | \text{Diskon}) \times P(\text{Pembayaran} = \text{ShopeePay/ShopeePayLater} | \text{Diskon})$   
 $= 0,046 \times 0,923 \times 0,923 \times 1,000 \times 1,000 \times 0,846 \times 0,923 \times 1,000 \times 0,462 \times 0,923$   
 $= 0,01304899$
5.  $P(\text{Tidak Ada} | X_1, X_2, X_3, X_n)$   
 $= P(\text{Tidak Ada}) \times P(\text{Usia} = \text{Usia} \leq 20 | \text{Tidak Ada}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Tidak Ada}) \times P(\text{Mengetahui Shopee} = \text{Ya} | \text{Tidak Ada}) \times P(\text{Memiliki Akun Shopee} = \text{Ya} | \text{Tidak Ada}) \times P(\text{Sebagai Apa Di Shopee} = \text{Pembeli} | \text{Tidak Ada}) \times P(\text{Rata-Rata Pembelian} = \text{Jarang} | \text{Tidak Ada}) \times P(\text{Menggunakan Promosi} = \text{Ya} | \text{Tidak Ada}) \times P(\text{Produk} = \text{Perawatan \& Kecantikan} | \text{Tidak Ada}) \times P(\text{Pembayaran} = \text{ShopeePay/ShopeePayLater} | \text{Tidak Ada})$   
 $= 0,042 \times 0,333 \times 0,583 \times 1,000 \times 1,000 \times 1,000 \times 1,000 \times 0,000 \times 0,250 \times 0,417$   
 $= 0$

3.1.4. Membandingkan Hasil Dari Setiap Kelas

Pada tabel 3 terdapat hasil perhitungan probabilitas dari contoh data testing, nilai yang sudah diperoleh selanjutnya dibandingkan dengan setiap nilai kelas. Pada metode Naïve Bayes nilai yang memiliki probabilitas tertinggi ialah label kelas yang diperoleh [17]. Dari data testing yang sudah diuji menghasilkan kelas (P | Gratis Ongkir) memiliki nilai probabilitas paling tinggi diantara kelas lainnya, sehingga dapat disimpulkan bahwa promosi gratis ongkir yang akan digunakan.

Tabel 3. Hasil Perhitungan Probabilitas Data Testing

Kelas	Probabilitas
(P   Gratis Ongkir)	0,09076653
(P   Flash Sale)	0,00056525

(P   Cashback)	0,00365349
(P   Diskon)	0,01304899
(P   Tidak Ada)	0

### 3.2. Evaluasi

Untuk mengukur model klasifikasi dengan metode Naïve Bayes dalam penelitian ini menggunakan *cross validation*. *Cross validation* bertugas membagi data kedalam ukuran yang sama untuk dilatih dan diuji dengan banyaknya subset [18]. Untuk mengevaluasi model klasifikasi dengan metode *cross validation* dibutuhkan skenario pengujian dengan melakukan pembagian data sesuai dengan jumlah *n-fold*. Tujuan utama dari penelitian ini adalah untuk mengetahui nilai akurasi dan kappa dari metode Naïve Bayes yang digunakan untuk mengklasifikasikan pengguna Shopee berdasarkan promosi. Pada tabel 4 merupakan hasil keseluruhan pengujian dengan menggunakan algoritma Naïve Bayes.

Tabel 4. Probabilitas Produk

Skenario	Accuracy	Precision	Recall	Kappa
2 fold	85,92	45,57	41,08	0,348
3 fold	88,73	64,42	45,41	0,451
4 fold	88,03	56,37	48,62	0,453
5 fold	87,32	53,43	46,04	0,42
6 fold	88,38	57,54	46,29	0,439
7 fold	87,68	63,45	44,54	0,415
8 fold	86,97	63,43	44,37	0,407
9 fold	88,03	64,4	46,2	0,406
10 fold	88,38	66,37	48,7	0,444

Dari hasil keseluruhan pengujian *cross validation* menggunakan algoritma Naïve Bayes didapatkan bahwa skenario dengan 2 fold memiliki nilai *accuracy* yang paling kecil yaitu 85,92 % dengan nilai *kappa* sebesar 0,348. Sedangkan skenario dengan 3 fold adalah skenario terbaik yang memiliki nilai *accuracy* tertinggi sebesar 88,73% dengan nilai *kappa* sebesar 0,451. Nilai *kappa* berapa pada kategori cukup (*moderate*) sehingga hasil klasifikasi pengguna Shopee menggunakan algoritma Naïve Bayes pada penelitian ini cukup konsisten [19].

Meskipun memiliki nilai *accuracy* yang tinggi tetapi nilai *kappa* yang dihasilkan tidak terlalu besar sehingga hanya masuk kedalam kategori cukup (*moderate*). Pada penelitian ini nilai *kappa* yang kecil dapat disebabkan karena adanya *imbalance* data. Data *imbalance* merupakan kondisi dimana suatu kelompok kelas memiliki jumlah data yang jauh berbeda jika dibandingkan dengan kelas lainnya. Data *mining* mengartikan *imbalance* dengan jumlah data kelas mayoritas memiliki lebih banyak data dibandingkan kelas minoritas [20].

## 4. SIMPULAN

Hasil dari penelitian yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Penelitian ini dilakukan untuk mengklasifikasikan pengguna Shopee berdasarkan promosi menggunakan algoritma *Naïve Bayes* dengan metodologi KDD. Dari hasil perhitungan klasifikasi menggunakan algoritma *Naïve Bayes* promosi gratis ongkir adalah promosi yang paling sering digunakan dengan pembelian produk *fashion* dan juga perawatan & kecantikan.
2. Hasil pengujian klasifikasi algoritma *Naïve Bayes* menghasilkan performa terbaik pada skenario pengujian dengan menggunakan 3 fold yang menghasilkan nilai *accuracy* sebesar 88,73% dan dengan nilai *kappa* yang diperoleh sebesar 0,451 yang termasuk kedalam kategori cukup (*moderate*). Berdasarkan hasil tersebut maka, model yang dihasilkan oleh algoritma *Naïve Bayes* ini cukup konsisten.



## 5. SARAN

Dari penelitian yang telah dilakukan, terdapat saran yang perlu disampaikan pada penelitian ini untuk referensi penelitian yang akan datang :

1. Menambahkan atribut yang lebih berpengaruh untuk klasifikasi pengguna Shopee agar akurasi algoritma Naïve Bayes semakin meningkat.
2. Pada penelitian ini perlu ditambahkan teknik SMOTE untuk mendapatkan performa klasifikasi yang lebih sempurna untuk menangani imbalance data atau data yang tidak seimbang.
3. Perlu ditambahkan skenario pengujian lainnya dalam membangun sebuah model untuk mendapatkan hasil yang lebih baik. Dapat menambahkan feature selection seperti forward selection dan backward selection, atau menggunakan algoritma klasifikasi lainnya.
4. Menggunakan E-commerce lain sebagai objek penelitian lainnya.

## DAFTAR PUSTAKA

- [1] D. A. Harahap, "Perilaku belanja online di Indonesia: studi kasus," *JRMSI - Jurnal Riset Manajemen Sains Indonesia*, vol. 9, no. 1, pp. 193-213, 2018.
- [2] N. R. Yunus dan A. Rezki, "Kebijakan Pemberlakuan Lock Down Sebagai Antisipasi Penyebaran Corona Virus Covid-19," *SALAM: Jurnal Sosial Dan Budaya Syar-I*, vol. 7, no. 227-238, p. 3, 2020.
- [3] G. Nurdian, "Graha Nurdian," 14 Februari 2021. [Online]. Available: <http://grahanurdian.com>. [Diakses 3 Maret 2021].
- [4] "Peta E-Commerce Indonesia," 2020. [Online]. Available: <https://iprice.co.id/insights/mapofecommerce/>.
- [5] Ardiyansyah and P. A. Rahayuningsih, "Penerapan teknik sampling untuk mengatasi imbalance class pada klasifikasi online shoppers intentionenri," *Jurnal Teknik Informatika Kaputama (JTIIK)*, vol. 4, no. 1, pp. 7-14, 2020.
- [6] A. M. Siregar dan A. Puspabhuana, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*, CV Kekata Group, 2017.
- [7] I. Ranggadara, G. Wang dan E. R. Kaburuan, "Applying Customer Loyalty Classification with RFM and Naïve Bayes for Better Decision Making," *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 564-568, 2019.
- [8] I. Romli, E. Pusnawati dan A. Siswandi, "PENENTUAN TINGKAT PENJUALAN MOBIL DI INDONESIA DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES," *e-Prosiding SNasTekS*, vol. 1, no. 1, pp. 367-380, 2019.
- [9] H. F. Putro, R. T. Vlandari dan W. L. Y. Saptomo, "Penerapan Metode Naïve Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, vol. 8, no. 2, 2020.
- [10] R. N. Devita, H. W. Herwanto and A. P. Wibawa, "Perbandingan kinerja metode naïve bayesian dan k-nearest neighbor untuk klasifikasi artikel berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 5, no. 4, pp. 427-434, 2018.
- [11] S. Rohmatunisha, U. Wahyudi and D. S. Yudasmara, "Survei minat siswa dalam mengikuti kegiatan ekstrakurikuler bolabasket pada peserta sekolah menengah pertama," *Sport Science and Health*, vol. 2, no. 2, pp. 119-129, 2020.
- [12] M. D. Purbolaksono, M. I. Tantowi, A. I. Hidayat dan A. , "Perbandingan Support Vector Machine dan Modified Balanced Random," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 393-399, 2021.
- [13] M. A. Muslim, E. Sugiharti, B. Prasetyo dan S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4. 5 dalam Mendiagnosa Diabetes," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 8, no. 2, pp. 135-143, 2017.

- [14] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*, Yogyakarta: Deepublish, 2014.
- [15] A. Mukminin dan D. Riana, "Komparasi Algoritma C4.5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah," *JURNAL INFORMATIKA*, vol. 4, no. 1, pp. 21-31, 2017.
- [16] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi," *JURNAL INFORMATIKA*, vol. 6, no. 1, pp. 18-22, 2019.
- [17] D. Sartika dan D. I. Sensus, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 3, no. 2, pp. 151-161, 2017.
- [18] S. Hulu, "Analisis kinerja metode cross validation dan k-nearest neighbor dalam klasifikasi data," Sumatera Utara, 2020.
- [19] U. Enri, "Penerapan Algoritma C4.5 Dalam Pemilihan Program Studi Fakultas Ilmu Komputer (Studi Kasus Sekolah Menengah Atas Negeri 1 Tambun Utara)," *URNAL REKAYASA INFORMASI*, vol. 7, no. 1, 2018.
- [20] A. M. WS, I. P. Ananda, M. A. Rizki, Z. D. Hapsari dan R. Nooraeni, "Penerapan Metode Resampling dalam Mengatasi Imbalanced Data Pada Determinan Kasus," *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 8, no. 1, pp. 19-27, 2020.