

# Komparasi Algoritma *Naive Bayes* dan *K Nearest Neighbor* dalam Kepuasan Pengguna Fitur Tiktok Shop

Balqist Naurah Salsabilah<sup>1</sup>, I Kadek Dwi Nuryana<sup>2</sup>

<sup>1,2</sup>Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya

[balqist.190830@mhs.unesa.ac.id](mailto:balqist.190830@mhs.unesa.ac.id)

[dwinuryana@unesa.ac.id](mailto:dwinuryana@unesa.ac.id)

**Abstrak**— Hadirnya teknologi yang semakin canggih membuat aktivitas masyarakat yang kini dilakukan dengan *online*, salah satunya transaksi jual beli. Minat penggunaan masyarakat di Indonesia dalam melakukan transaksi jual beli online meningkat dari tahun ke tahun. Berbagai platform *e-commerce* memenuhi pasar Indonesia, salah satunya fitur pada aplikasi Tiktok. Pada tahun 2021, Shopify dan Tiktok mengenalkan fitur Tiktok Shop sebagai *social commerce*. Namun, masih banyak pengguna yang memberikan *review* ataupun *rating* yang kurang pada *Google Play Store* serta *App Store* mengenai fitur Tiktok Shop.

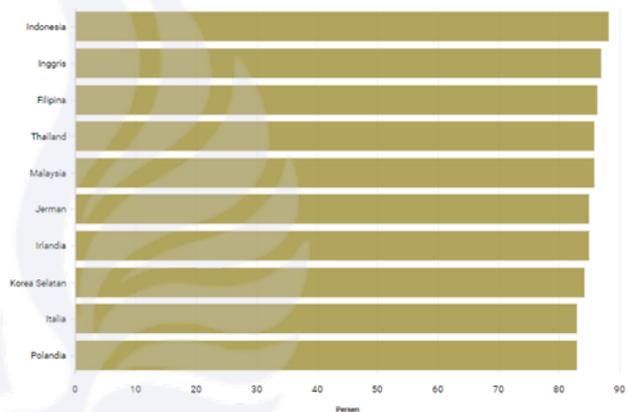
Fenomena ini menarik perhatian peneliti untuk mengetahui faktor apa saja yang mempengaruhi kepuasan pelanggan terhadap fitur Tiktok Shop. Penilaian kepuasan pengguna diperlukan untuk mengetahui apakah sudah memenuhi ekspektasi dari para pengguna atau belum. Indikator penilaian kepuasan pengguna berdasarkan konstruk model *DeLone and McLean IS Success*. Penelitian ini akan menggunakan algoritma *Naive Bayes* dan *K Nearest Neighbor* karena hasil tingkat akurasi yang pernah dilakukan dalam melakukan klasifikasi data memberikan hasil yang baik dan akan dibandingkan. Penelitian ini juga akan mencari tahu faktor-faktor apa saja yang memiliki pengaruh tinggi terhadap kepuasan pengguna fitur Tiktok Shop pada pada aplikasi Tiktok. Dalam mencari tahu faktor-faktor yang memiliki pengaruh yakni menggunakan *Gain Ratio* dan mengurutkannya berdasarkan yang paling tinggi hingga terendah. Hasil yang didapatkan yakni *service quality* menjadi pengaruh paling tinggi pada kepuasan pengguna fitur Tiktok Shop, selanjutnya variabel *system quality*, dan variabel pada urutan terakhir adalah *information quality*. Hasil dari perbandingan algoritma, *Naive Bayes* memiliki bobot akurasi terbaik sebesar 98.36% dengan 70% pelatihan data dan 30% pengujian data.

**Kata Kunci**— *Tiktok Shop, Naive Bayes, K Nearest Neighbor, Kepuasan Pengguna, Algoritma.*

## I. PENDAHULUAN

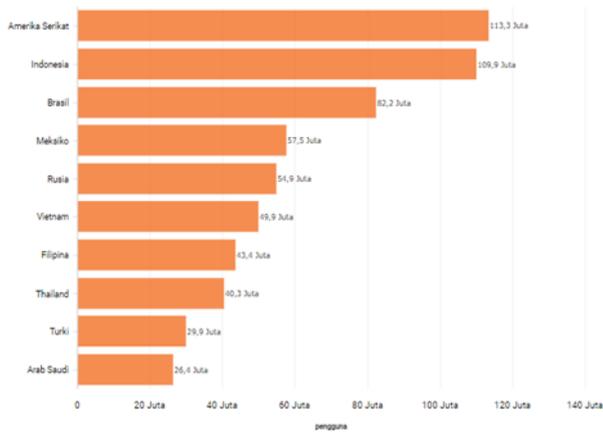
Hadirnya teknologi yang semakin canggih membuat aktivitas masyarakat yang kini dilakukan dengan *online*, salah satunya transaksi jual beli. Transaksi jual beli secara *online* memudahkan para pelaku usaha untuk menjual dan mempromosikan produk miliknya, juga memudahkan pembeli dalam membeli produk yang dibutuhkan. Minat penggunaan masyarakat di Indonesia dalam melakukan transaksi jual beli *online* meningkat dari tahun ke tahun. Potensi tersebut tentu tidak dilewatkan oleh para pengembang aplikasi untuk mengoptimalkan aplikasi agar meningkatkan rasa minat pengguna dalam menggunakan aplikasi mereka. Adanya platform *e-commerce* ini untuk melakukan transaksi jual beli, sangat memudahkan para penggunanya melakukan pembelian barang dengan praktis tanpa perlu mendatangi toko.

Produk yang hadir pada platform *e-commerce* online pun beragam macam. Dari barang kebutuhan sehari-hari, sembako, fashion, produk kecantikan, bahkan alat elektronik. Pada tahun 2021, berdasarkan data dari databoks, Indonesia menjadi negara urutan pertama dengan presentase pengguna *e-commerce* sebanyak 88,1% [12].



Gbr. 1 Negara dengan Persentase Penggunaan E-Commerce Tertinggi di Dunia

Berbagai platform *e-commerce* memenuhi pasar Indonesia, salah satunya Tiktok. Tiktok merupakan aplikasi sosial media yang sedang populer di berbagai negara, terlebih di Indonesia. Berdasarkan data grafik dari katadata.co.id, Indonesia menduduki urutan kedua sebagai negara pengguna terbanyak aplikasi Tiktok setelah di bawah Amerika Serikat. Menurut databoks, pengguna Tiktok dari berbagai usia di Indonesia mencapai hampir 109.09 juta pada tahun 2023 dan akan terus bertambah [4].



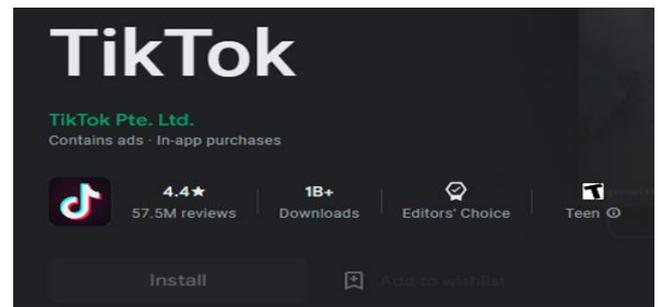
Gbr. 2 Negara dengan Jumlah Pengguna TikTok Terbanyak di Dunia

Karena TikTok termasuk sebagai aplikasi dengan pengguna dan pengunduh yang terus bertambah, TikTok berinovasi dengan menghadirkan fitur-fitur baru seperti TikTok Shop. Pada tahun 2021, Shopify dan TikTok mengenalkan fitur TikTok Shop sebagai *social commerce*. Meskipun TikTok Shop bukan benar-benar *e-commerce*, namun cara penggunaan, alur pembelian, dan lainnya sama seperti *e-commerce* pada umumnya. Pengguna dapat memasarkan produk mereka dan pembeli dapat melakukan pembelian langsung. Hasil dari survei yang dilakukan Populix dengan 880 orang mengenai transaksi jual beli *online*, platform yang paling banyak digunakan adalah TikTok Shop sebanyak 46%, untuk urutan kedua diduduki Whatsapp, dan Facebook Shop ada pada urutan terakhir [15].



Gbr. 3 Penggunaan Social Media untuk Berbelanja

Hal tersebut menunjukkan bahwa aplikasi TikTok memiliki peran penting bagi para pelaku usaha untuk melakukan operasi komersial transaksi jual beli dengan pembeli. Manfaat dari TikTok Shop dapat dirasakan langsung oleh para pelaku usaha. Mereka dapat menjual sekaligus mempromosikan produk-produk mereka. Namun, masih banyak pengguna yang memberikan *review* ataupun *rating* yang kurang pada Google Play Store. *Rating* pada Google Play Store yakni sebesar 4.4 dari 57.5 juta pengguna yang memberikan *rating*.



Gbr. 4 Rating Tiktok Shop

Fenomena ini menarik perhatian peneliti untuk mengetahui faktor apa saja yang mempengaruhi kepuasan pelanggan terhadap fitur Tiktok Shop. Penilaian kepuasan pengguna diperlukan untuk mengetahui apakah sudah memenuhi ekspektasi dari para pengguna atau belum. Indikator penilaian kepuasan pengguna berdasarkan konstruk model *DeLone and McLean IS Success*. Model ini menyebutkan jika *Information Quality*, *System Quality*, *Service Quality* akan berpengaruh positif terhadap *User Satisfaction*. DeLone dan McLean sebagai salah satu model yang biasa digunakan di berbagai bidang selama bertahun-tahun. Beberapa bidang yang telah dimanfaatkan antara lain penilaian sistem *e-commerce*, sistem manajemen, dan *e-government*. Kepuasan pengguna dari suatu sistem pada DeLone and McLean dapat digunakan menjadi salah satu pengukur keberhasilan suatu sistem informasi.

Model dari DeLone and McLean sudah banyak digunakan oleh peneliti sebelumnya untuk mengetahui tingkat kepuasan pengguna terhadap sistem informasi. Penelitian yang menggunakan model DeLone and McLean dalam mengukur kepuasan pengguna salah satunya dilakukan oleh Didik dan Triyonowati (2022), hasil yang didapat yakni *System Quality*, *Information Quality*, dan *Service Quality* berpengaruh terhadap *User Satisfaction* [23]. Berdasarkan pemaparan penelitian terdahulu yang menghasilkan bahwa model dari DeLone and McLean dapat digunakan untuk melakukan pengukuran kepuasan pengguna.

Kemudian pada penelitian yang dilakukan oleh Yusran T. Samuel dan Kemala Dewi (2019) dengan judul Penggunaan Metode *NAIVE BAYES* Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent Indonesia yang melakukan pengukuran kepuasan pengguna menggunakan algoritma *Naive Bayes* dengan akurasi tertinggi yakni sebesar 81.3% menggunakan *full data training*. Dari 8 atribut yang digunakan, 5 atribut yang memiliki pengaruh terhadap tingkat kepuasan pengguna dengan nilai *gain ratio* yang paling tinggi yakni Kualitas Konten, Kecepatan, Daya Tarik Visual, Kelengkapan Secara *Online*, dan *Customer Service*. 3 atribut dengan nilai *gain ratio* yang rendah adalah *Relevan*, *Privasi*, dan Mudah Dioperasikan [20].

Pada penelitian yang dilakukan oleh Jirapon dan Markus (2019) dengan melakukan perbandingan tiga algoritma, yakni *Decision Tree*, *Naive Bayes*, dan *Logistic Regression*. Hasil akurasi yang didapat untuk *Naive Bayes* mencapai 92.76%, 91.03% untuk *Logistic Regression* dan 92.41% untuk *Decision*

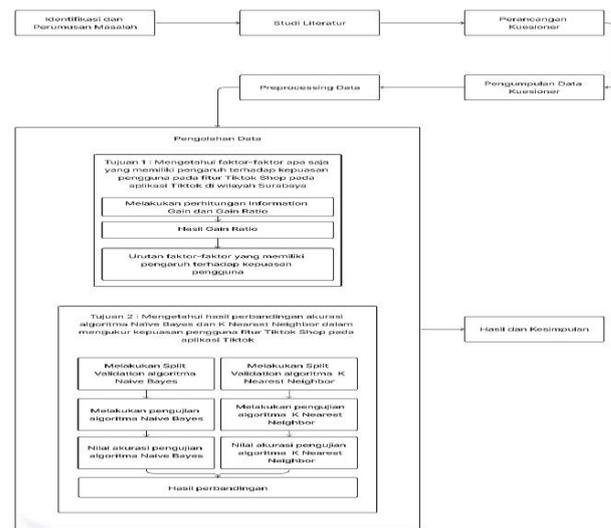
Tree. Lalu pada penelitian yang dilakukan Herdi, et al (2023) dengan membandingkan algoritma *Support Vector Machine* dan *Naive Bayes* untuk melakukan klasifikasi, akurasi yang didapatkan algoritma *Naive Bayes* sebesar 97,39% dan 96,52% untuk *Support Vector Machine*. Hasil dari kedua penelitian tersebut menyatakan bahwa *Naive Bayes* memiliki tingkat akurasi yang lebih tinggi dalam melakukan klasifikasi [11].

Kemudian penelitian yang dilakukan oleh Devi Yunita (2017) membandingkan algoritma *K Nearest Neighbor* dan Decision Tree untuk klasifikasi penentuan risiko kredit kepemilikan mobil, akurasi yang didapatkan untuk K-Nearest Neighbor sebesar 98,18 %, sedangkan untuk Decision Tree sebesar 90,91% [29].

Penelitian ini akan menggunakan algoritma *Naive Bayes* dan karena hasil tingkat akurasi dalam melakukan klasifikasi data memberikan hasil yang baik dan kemudian hasil dari pengujian data menggunakan algoritma *Naive Bayes* akan dibandingkan dengan algoritma *K Nearest Neighbor*. Dengan membandingkan kedua algoritma ini, akan membantu untuk menentukan algoritma mana yang memiliki akurasi dalam melakukan klasifikasi kelas puas dan tidak puas. Menurut Lantz (dalam Dela, 2021), algoritma *Naive Bayes* merupakan algoritma dapat berfungsi baik dengan rasio data latih yang kecil. Begitu pula untuk algoritma *K Nearest Neighbor*, dapat berfungsi baik dengan data yang kecil karena tidak membutuhkan pelatihan [22]. Penelitian ini juga akan mencari tahu faktor-faktor apa saja yang memiliki pengaruh tinggi terhadap kepuasan pengguna fitur Tiktok Shop pada pada aplikasi Tiktok. Berdasarkan penjelasan yang dipaparkan, penulis memutuskan untuk melakukan penelitian dengan judul “Komparasi Akurasi Algoritma *Naive Bayes* dan *K Nearest Neighbor* dalam Klasifikasi Kepuasan Pengguna Fitur Tiktok Shop pada Aplikasi Tiktok”.

## II. METODOLOGI PENELITIAN

Rancangan penelitian merupakan langkah-langkah yang terstruktur berisi proses dilakukan dalam penelitian dari tahap awal hingga pada tahap akhir. Adapun langkah-langkah yang dilakukan peneliti pada penelitian ini tertera pada gambar 5.



Gbr. 5 Metodologi Penelitian

### A. Identifikasi dan Perumusan Masalah

Pada tahap ini dilakukan identifikasi masalah yang kemudian digunakan untuk menyusun perumusan masalah. Rumusan masalah yang ada pada penelitian ini didapat dari hasil observasi terhadap fenomena.

### B. Studi Literatur

Pada tahapan ini, peneliti melakukan studi literatur untuk mendapatkan referensi yang sesuai pada objek penelitian ini. Studi literatur dilakukan dengan mengumpulkan, memahami, serta mempelajari literatur dari buku, jurnal, artikel, maupun penelitian lainnya.

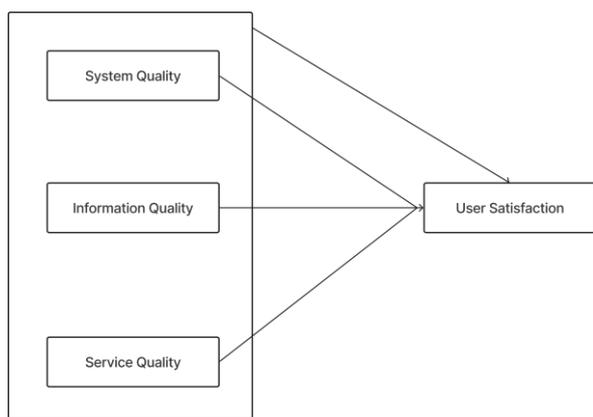
### C. Membuat Model Konseptual

Digunakan model konseptual untuk memberikan penjelasan tentang gagasan penelitian yang akan menjadi dasar kerangka penelitian.

### D. Perancangan Kuesioner

Pada tahapan ini, peneliti melakukan penentuan kriteria responden sebelum merancang kuesioner. Kriteria responden yang digunakan pada penelitian ini adalah pengguna aplikasi Tiktok berumur minimal 17 tahun yang pernah melakukan pembelian dan menggunakan fitur Tiktok Shop di wilayah Kota Surabaya. Penelitian ini merupakan *data collection*, yang dimana data yang dibutuhkan pada penelitian ini adalah sebanyak-banyaknya. Jumlah sampel tidak ditetapkan pada penelitian ini.

Penyusunan kuesioner berdasarkan indikator dari konstruk model DeLone and McLean IS Success yang kemudian diajukan sebagai pernyataan untuk diberikan kepada responden.



Gbr. 6 Model Konseptual

Instrumen dari penelitian ini yang akan digunakan pada yakni menggunakan kuesioner dari indikator konstruk model DeLone and McLean IS Success untuk mengukur kepuasan pengguna terhadap fitur Tiktok Shop menggunakan algoritma *Naive Bayes*. Berikut merupakan instrumen penelitian yang digunakan penelitian ini:

TABEL I  
INSTRUMEN PENELITIAN (WILLIAM H. DELONE DAN EPHRAIM R. MCLEAN)

Variabel	Kode	Indikator
System Quality	SQ1	Adaptability (Adaptasi)
	SQ2	Availability (Ketersediaan)
	SQ3	Reliability (Keandalan)
	SQ4	Response time (Waktu respon)
	SQ5	Usability (Kegunaan)
Information Quality	IQ1	Completeness (Kelengkapan)
	IQ2	Ease of understanding (Kemudahan pemahaman)
	IQ3	Personalization (Personalisasi)
	IQ4	Relevance (Relevan)
	IQ5	Accuracy (Keakuratan)
Service Quality	SEQ1	Assurance (Jaminan)
	SEQ2	Empathy (Empati)
	SEQ3	Responsiveness (Kecepatan respon)
User Satisfaction	US	Repeat visits (Kunjungan ulang)

Skala Likert jenis skala peringkat yang sering ditemukan pada survei atau kuesioner dan digunakan untuk mengukur pendapat orang terhadap sesuatu. Skala Likert dapat berguna dalam berbagai keadaan. Pembobotan pada skala Likert yang akan digunakan pada penelitian ini yakni sebagai berikut.

TABEL II  
SKALA LIKERT

No	Keterangan	Nilai
1	Sangat Tidak Setuju	1
2	Tidak Setuju	2
3	Netral	3
4	Setuju	4
5	Sangat Setuju	5

E. Pengumpulan Data Kuesioner

Pada tahapan ini, peneliti melakukan pengumpulan data kuesioner setelah melakukan penyusunan pernyataan yang diajukan untuk responden. Pengumpulan data kuesioner akan disebar melalui Google Form. Data kuesioner akan diisi oleh responden yang telah ditentukan pada ketika melakukan sampling populasi.

F. Preprocessing Data

Preprocessing data yakni melakukan menghilangkan data yang salah dan menghapus data yang kolomnya kosong. Selanjutnya penilaian dari skala likert dibagi menjadi 2 interval nilai untuk mengklasifikasikan hasil kuesioner menjadi 2 kelas, yakni kelas puas dan tidak puas [5]. Penilaian 2 interval ada pada tabel berikut:

TABEL III  
INTERVAL NILAI

Interval Nilai	Klasifikasi
1-3	Tidak Puas
4-5	Puas

G. Analisis Data

Dalam mengetahui faktor-faktor apa saja yang memiliki pengaruh paling tinggi terhadap kepuasan pengguna dilakukan perhitungan menggunakan *gain ratio*. Teknik *gain ratio* meningkatkan pendekatan *information gain* dengan menggabungkan informasi intrinsik dari setiap atribut. *Gain Ratio* memilih atribut yang optimal dengan mengevaluasi nilai *gain ratio* yang paling tinggi. Perhitungan untuk *Gain Ratio* dapat dihitung dengan menggunakan rumus:

$$Gain\ Ratio(A) = \frac{Gain(A)}{SplitInfo(A)} \tag{1}$$

Keterangan:  
*Gain(A)* = Perhitungan Information Gain  
*SplitInfo(A)* = Split perhitungan entropy

Rumus dari *Information Gain* adalah sebagai berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{Si}{S} \right| x Entropy(Si) \tag{2}$$

Keterangan:  
*A* = Atribut  
*S* = Sampel  
*|Si|* = Jumlah dari sampel untuk nilai *i*

$|S|$  = Jumlah untuk seluruh sampel data  
 $Entropy(S)$  = nilai *Entropy* Sampel

Dalam mencari hasil dari *Information Gain*, diperlukan rumus entropy dengan persamaan sebagai berikut:

$$Entropy(S) = \sum_i^c - p_i \log_2 p_i \quad (3)$$

Keterangan:  
 $c$  = Jumlah kelas klasifikasi  
 $p_i$  = Porsi sampel kelas  $i$

Kemudian rumus untuk *SplitInfo(A)* yang akan digunakan untuk mencari *Gain Ratio* dengan persamaan sebagai berikut:

$$SplitInfo_A^{(S)} = - \sum_{i=1}^n \left| \frac{S_i}{S} \right| x \log \log_2 \left| \frac{S_i}{S} \right| \quad (4)$$

Keterangan:  
 $A$  = Atribut  
 $|S_i|$  = Jumlah dari sampel untuk nilai  $i$   
 $|S|$  = Jumlah untuk seluruh sampel data

Dalam melakukan analisis data dengan *Naive Bayes* juga *K Nearest Neighbor*, *split validation* digunakan sebagai pengujian data dengan membagi dataset yang ada menjadi dua yaitu data *training* dan data *test*. Distribusi data *training* serta data *testing* dilakukan sebelum melakukan pengukuran dengan *Naive Bayes* dan *K Nearest Neighbor*. Akan ada total tiga pengujian data yang dilakukan, pengujian pertama akan melibatkan 80% pelatihan data dan 20% pengujian data. Pengujian kedua akan melibatkan 70% pelatihan data dan 30% pengujian data. Pengujian ketiga 90% pelatihan data dan 10% pengujian data.

Nilai probabilitas yang digunakan pada *Naive Bayes Classifier* untuk menyatakan secara sederhana berdasarkan pada terema Bayes adalah sebagai berikut:

$$p(Y|Xn) = \frac{p(Xn|Y) p(Y)}{p(Xn)} \quad (5)$$

Keterangan:  
 $Xn$  = Data sample  
 $Y$  = Hipotesa  
 $p(Y|Xn)$  = Probabilitas hipotesis  $Y$  untuk data sampel  $Xn$  yang diamati  
 $p(Xn|Y)$  = Probabilitas  $Xn$ , berdasarkan hipotesis  $Y$   
 $p(Y)$  = Probabilitas hipotesis  $Y$   
 $p(Xn)$  = Probabilitas hipotesis  $Xn$

Dalam *K Nearest Neighbor* mengukur jarak antara data training dengan data testing. Berikut merupakan persamaan dari rumus *Euclidean Distance* yang digunakan untuk mengukur jarak antara data latih dengan data uji:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

Keterangan:  
 $d(x, y)$  = Jarak antara data  $x$  ke data  $y$   
 $n$  = Jumlah fitur  
 $i$  = Jumlah data uji ke-  
 $x$  = Data uji  
 $y$  = Data latih

Lalu dilakukan evaluasi untuk performa dengan menggunakan *Confusion Matrix*. Terdapat 3 nilai kombinasi untuk melakukan perhitungan *Confusion Matrix*, yaitu *accuracy*, *precision*, dan *recall*.

Rumus untuk melakukan perhitungan akurasi adalah sebagai berikut:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Rumus untuk melakukan perhitungan *precision* adalah sebagai berikut:

$$precision = \frac{TP}{TP+FP} \quad (8)$$

Rumus untuk melakukan perhitungan *recall* adalah sebagai berikut:

$$recall = \frac{TP}{TP+FN} \quad (9)$$

#### H. Hasil Penelitian

Hasil penelitian merupakan hasil akhir atas tahapan-tahapan yang dilakukan sebelumnya. Hasil bisa berupa kesimpulan ataupun data yang didapat setelah dianalisis. Kesimpulan atau data yang didapat berdasarkan hasil analisis yang dilakukan sebelumnya.

### III. HASIL DAN PEMBAHASAN

Bagian ini akan membahas mengenai data kuesioner yang didapat dan pengolahan data awal dengan melakukan uji validitas dan realibilitas. Setelah melakukan dua uji tersebut, data akan diolah dengan *RapidMiner* untuk mencari tahu akurasi dari penggunaan algoritma *Naive Bayes* dalam mengukur tidak puas ataupun puasnya pengguna dalam menggunakan fitur *Tiktok Shop*.

#### A. Uji Validitas dan Realibilitas

Yang pertama yakni pengujian validitas yang dilakukan untuk membandingkan hasil dari nilai antara nilai  $r$  hitung dengan nilai pada  $r$  table. Instrumen dapat dinyatakan valid apabila  $r$  hitung lebih besar daripada  $r$  tabel. Untuk melakukan responden sebanyak 34 responden. Berdasarkan pada nilai  $r$ -tabel, nilai untuk 34 responden dengan level signifikansi sebesar 5% adalah 0.399.

TABEL IV  
UJI VALIDITAS

No	Variabel	R Hitung	R Tabel	Ket
1	SQ1	0.761	0.339	Valid

2	SQ2	0.764	0.339	Valid
3	SQ3	0.782	0.339	Valid
4	SQ4	0.724	0.339	Valid
5	SQ5	0.795	0.339	Valid
6	IQ1	0.693	0.339	Valid
7	IQ2	0.822	0.339	Valid
8	IQ3	0.803	0.339	Valid
9	IQ4	0.861	0.339	Valid
10	IQ5	0.877	0.339	Valid
11	SEQ1	0.770	0.339	Valid
12	SEQ2	0.871	0.339	Valid
13	SEQ3	0.836	0.339	Valid

Uji realibilitas digunakan untuk mengetahui dan menentukan konsisten tidaknya instrumen apabila digunakan berkali-kali. Pengujian dilakukan untuk mengetahui nilai dari Cronbach's Alpha, apabila nilai dari Cronbach's Alpha lebih dari 0.6 maka perhitungan uji reliabilitas diterima.

TABEL V  
UJI REALIBILITAS

No	Variabel	Cronbach's Alpha	Ket
1	SQ	0.800	Reliabel
2	IQ	0.865	Reliabel
3	SEQ	0.760	Reliabel

### B. Analisa Information Gain dan Gain Ratio

Information Gain dilakukan untuk mengetahui apa saja atribut-atribut yang memiliki pengaruh paling tinggi dari indikator pengukuran kepuasan pengguna. Nilai Information Gain diurutkan dari nilai tinggi hingga rendah yang berarti nilai tertinggi adalah atribut paling berpengaruh.

TABEL VI  
HASIL INFORMATION GAIN

Urutan	Atribut		Information Gain
1	Responsiveness	SEQ1	0.171
2	Ease of Understanding	IQ1	0.168
3	Empathy	SEQ2	0.154
4	Adapatability	SQ1	0.148
5	Availability	SQ3	0.113
6	Usability	SQ2	0.108
7	Accuracy	IQ2	0.102
8	Completeness	IQ4	0.091
9	Personalization	IQ3	0.089
11	Assurance	SEQ3	0.075
11	Relevance	IQ5	0.072
12	Response Time	SQ4	0.071
13	Realibility	SQ5	0.024

Gain ratio adalah perkembangan dari information gain yang dapat digunakan untuk menurunkan bias. Gain Ratio akan

dilakukan untuk mengetahui apa saja atribut-atribut yang berpengaruh terhadap pengukuran kepuasan pengguna. Nilai gain ratio diurutkan dari nilai tinggi hingga rendah yang berarti nilai tertinggi adalah atribut paling berpengaruh.

TABEL VII  
HASIL GAIN RATIO

Urutan	Atribut		Gain Ratio
1	Responsiveness	SEQ1	0.103
2	Ease of Understanding	IQ1	0.098
3	Empathy	SEQ2	0.093
4	Adaptability	SQ1	0.088
5	Usability	SQ2	0.075
6	Availability	SQ3	0.073
7	Accuracy	IQ2	0.057
8	Personalization	IQ3	0.049
9	Completeness	IQ4	0.048
10	Assurance	SEQ3	0.042
11	Relevance	IQ5	0.042
12	Response Time	SQ4	0.039
13	Realibility	SQ5	0.011

TABEL VIII  
RATA-RATA GAIN RATIO

No	Variabel	Atribut	Gain Ratio	Rata-rata
1	System Quality	Adaptability	0.088	0.062
		Usability	0.075	
		Availability	0.073	
		Adaptability	0.088	
		Response Time	0.039	
Realibility	0.011			
2	Information Quality	Ease of Understanding	0.098	0.059
		Accuracy	0.057	
		Personalization	0.049	
		Completeness	0.048	
		Relevance	0.042	
3	Service Quality	Responsiveness	0.103	0.079
		Empathy	0.093	
		Assurance	0.042	

Berdasarkan pada hasil gain ratio di atas, atribut yang memiliki nilai gain ratio paling tinggi adalah Responsiveness, pengguna merasa sistem dari fitur Tiktok Shop memberikan respon yang cepat sesuai kebutuhan. Atribut Realibility memiliki nilai yang paling rendah yakni sebesar 0.011. Pengguna merasa bahwa fitur Tiktok Shop masih kurang bisa diandalkan.

### C. Pengujian Algoritma

Akan dilakukan 3 kali pengujian menggunakan split validation. Pengujian data dengan split validation dilakukan pembagian data menjadi dua, yakni menjadi data latihan dan data

uji. Pengujian pertama melibatkan 80% pelatihan data dan 20% pengujian data, yang kedua akan melibatkan 70% pelatihan data dan 30% pengujian data, dan terakhir 90% pelatihan data dan 10% pengujian data. Pengujian data dilakukan 3 kali untuk mengetahui pengujian mana yang memiliki tingkat akurasi paling tinggi.

1) Naive Bayes

Dilakukan beberapa pengujian menggunakan algoritma Naive Bayes dengan tools RapidMiner

- a. Pengujian pertama dilakukan dengan algoritma Naive Bayes pada RapidMiner menggunakan 80% pelatihan data dan 20% pengujian data dengan hasil sebagai berikut.

accuracy: 96.00%			
	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	9	2	81.82%
pred. PUAS	0	39	100.00%
class recall	100.00%	95.12%	

Gbr. 7 Hasil Naive Bayes 80% pelatihan data dan 20% pengujian data

Tingkat akurasi yang didapatkan sebanyak 96%. Berdasarkan gambar di atas, prediksi puas dan benar puas sebanyak 39, prediksi puas dengan kenyataan tidak puas sebanyak 0, prediksi tidak puas dengan kenyataan puas adalah 2, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 9. Hasil recall yang didapatkan sebesar 95.12% dan sebesar 100% untuk precision.

- b. Pengujian kedua dilakukan dengan membagi data menjadi 70% pelatihan data dan 30% pengujian data. Hasil pengujian kedua terdapat pada gambar di bawah.

accuracy: 98.65%			
	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	13	1	92.86%
pred. PUAS	0	60	100.00%
class recall	100.00%	98.36%	

Gbr. 8 Hasil Naive Bayes 70% pelatihan data dan 30% pengujian data

Tingkat akurasi yang didapatkan sebanyak 98.65%. Berdasarkan gambar di atas, hasil dari 70% pelatihan data dan 30% pengujian data adalah prediksi puas dan benar puas sebanyak 60, prediksi puas dengan kenyataan tidak puas sebanyak 0, prediksi tidak puas dengan kenyataan puas adalah 1, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 13. Hasil untuk precision sebesar 100% dan 98.36% untuk recall.

- c. Pengujian ketiga dengan 90% pelatihan data dan 10% pengujian data mendapatkan hasil sebagai berikut:

accuracy: 95.83%			
	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	4	1	80.00%
pred. PUAS	0	19	100.00%
class recall	100.00%	95.00%	

Gbr. 9 Hasil Naive Bayes 90% pelatihan data dan 10% pengujian data

Tingkat akurasi yang didapatkan sebanyak 95.83%. Pengujian data ini memiliki tingkat akurasi yang paling tinggi. Berdasarkan gambar di atas, hasil yang didapatkan adalah prediksi puas dan benar puas sebanyak 19, prediksi puas dengan kenyataan tidak puas sebanyak 0, prediksi tidak puas dengan kenyataan puas adalah 1, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 4. Hasil precision 100% dan recall 95%.

2) K Nearest Neighbor

Dilakukan beberapa pengujian menggunakan algoritma K Nearest Neighbor dengan tools RapidMiner

- a. Pengujian pertama dilakukan dengan 80% pelatihan data dan 20% pengujian data. Hasil yang didapatkan pada pengujian ini adalah sebagai berikut.

accuracy: 92.00%			
	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	5	0	100.00%
pred. PUAS	4	41	91.11%
class recall	55.56%	100.00%	

Gbr. 10 Hasil KNN 80% pelatihan data dan 20% pengujian data

Tingkat akurasi yang didapatkan sebanyak 92%. Berdasarkan gambar di atas, prediksi puas dan benar puas sebanyak 41, prediksi puas dengan kenyataan tidak puas sebanyak 4, prediksi tidak puas dengan kenyataan puas adalah 0, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 5. Hasil untuk precision sebesar 91.1% dan 100% untuk recall.

- b. Pengujian kedua dilakukan dengan membagi data menjadi 70% pelatihan data dan 30% pengujian data. Hasil pengujian kedua terdapat pada gambar di bawah.

accuracy: 90.54%			
	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	6	0	100.00%
pred. PUAS	7	61	89.71%
class recall	46.15%	100.00%	

Gbr. 11 Hasil KNN 70% pelatihan data dan 30% pengujian data

Tingkat akurasi yang didapatkan sebanyak 90.54%. Berdasarkan gambar di atas, prediksi puas dan benar puas sebanyak 61, prediksi puas dengan kenyataan tidak puas sebanyak 7, prediksi tidak puas dengan kenyataan puas adalah 0, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 6. Hasil untuk precision sebesar 89.71% dan 100% untuk recall.

- c. Pengujian ketiga dilakukan dengan membagi data menjadi 90% pelatihan data dan 10% pengujian data. Hasil pengujian kedua terdapat pada gambar di bawah.

accuracy: 91.67%

	true TIDAK PUAS	true PUAS	class precision
pred TIDAK PUAS	2	0	100.00%
pred PUAS	2	20	90.91%
class recall	50.00%	100.00%	

Gbr. 12 Hasil KNN 90% pelatihan data dan 10% pengujian data

Tingkat akurasi yang didapatkan adalah 91.67%. Berdasarkan gambar di atas, prediksi puas dan benar puas sebanyak 20, prediksi puas dengan kenyataan tidak puas sebanyak 2, prediksi tidak puas dengan kenyataan puas adalah 0, dan prediksi tidak puas dengan kenyataan benar tidak puas sebanyak 2. Hasil untuk *precision* sebesar 90.91% dan 100% untuk *recall*.

#### D. Hasil Perbandingan

Berdasarkan hasil dari pengujian menggunakan algoritma *Naive Bayes* dan *K Nearest Neighbor*, algoritma *Naive Bayes* memiliki nilai akurasi paling tinggi, yakni sebesar 100%. Sedangkan untuk algoritma *K Nearest Neighbor* memiliki nilai akurasi 91.67%

TABEL IX  
 HASIL PERBANDINGAN UJI ALGORITMA

Algoritma	Data latih	Data Uji	Accur acy	Precis ion	Rec all
<i>Naive Bayes</i>	90%	10%	95.83 %	100%	95%
	80%	20%	96%	100%	95.12%
	70%	30%	98.65 %	100%	98.36%
<i>K Nearest Neighbor</i>	90%	10%	91.67 %	90.91 %	100 %
	80%	20%	92%	91.2 %	100 %
	70%	30%	90.54 %	89.71 %	100 %

Dari hasil perbandingan, algoritma *Naive Bayes* lebih unggul mencapai akurasi 100% menggunakan *split validation* 90% dengan pelatihan data dan 10% pengujian data. Hasil tertinggi yang didapat dari penggunaan algoritma *K Nearest Neighbor* adalah sebesar 91.67% menggunakan *split validation* dengan 90% pelatihan data dan 10% pengujian data.

#### IV. KESIMPULAN

Berdasarkan pembahasan dan hasil dari penelitian yang telah dilakukan dalam mengukur tingkat kepuasan pengguna fitur Tiktok Shop dengan menggunakan algoritma *Naive Bayes*, dapat ditarik sebuah kesimpulan sebagai berikut:

- a. Berdasarkan hasil yang didapat dari *gain ratio*, atribut diurutkan dari yang memiliki pengaruh paling tinggi hingga paling rendah yaitu *Responsiveness*, *Ease of Understanding*, *Empathy*, *Adaptability*, *Usability*, *Availability*, *Accuracy*, *Personalization*, *Completeness*, *Assurance*, *Relevance*, *Response Time*, dan *Reliability*. Atribut yang menjadi faktor kepuasan pelanggan paling tinggi yakni *Responsiveness* dan atribut *Reliability* menjadi faktor yang paling rendah. *Service quality* menjadi pengaruh paling tinggi pada kepuasan pengguna fitur Tiktok Shop, selanjutnya variabel *system quality*, dan variabel pada urutan terakhir adalah *information quality*. *Service quality* yang diberikan Tiktok Shop memiliki nilai paling tinggi, namun diperlukan perbaikan dari sisi aspek *assurance* untuk meningkatkan pelayanan yang ada dalam memberikan rasa aman agar tidak ada rasa ragu-ragu dalam memakai fitur Tiktok Shop bagi para pengguna dengan menanamkan rasa percaya pelanggan. Selain itu memberikan solusi atas permasalahan atau complain dari para pengguna dengan cepat dan tanggap. *System quality* yang ada pada Tiktok Shop berada pada urutan kedua. Hal yang perlu diperhatikan adalah *reliability*. Keandalan dari sistem masih kurang, perlu ditingkatkan lagi ketahanan sistem yang ada dari kerusakan, kelambatan, maupun kesalahan dalam menanggapi permintaan pengguna. Waktu respon sistem atau *loading* dalam menanggapi permintaan dari pengguna juga perlu ditingkatkan. Layanan harus memberikan data ataupun informasi valid dan akurat serta menyediakan data tambahan atau kemungkinan sesuai permintaan pengguna. *Information quality* yang diberikan perlu ditingkatkan lagi dengan memberikan informasi yang relevan, akurat dan lengkap kepada pengguna. Ketidaklengkapan suatu informasi dapat mengakibatkan kekeliruan informasi, persepsi dan keputusan yang salah. Perlu disajikan informasi yang jelas, tidak ada kesalahan data maupun error. Informasi yang relevan berarti relevan terhadap apa yang pengguna minta atau inginkan dan apakah pengguna tersebut benar benar membutuhkan informasi tersebut.
- b. Berdasarkan pada hasil yang didapatkan setelah melakukan uji coba *split validation* menggunakan algoritma *Naive Bayes* sebanyak tiga kali uji coba, pada pengujian pertama menggunakan 90% pelatihan data dan 10% pengujian data memperoleh akurasi sebesar 95.83%. Pengujian kedua menggunakan 80% pelatihan data dan 20% pengujian data memperoleh akurasi sebesar 96%. Pada pengujian terakhir menggunakan 70% pelatihan data dan 30% pengujian data sebesar 98.65%. Berdasarkan pada hasil yang

didapatkan setelah melakukan uji coba *split validation* menggunakan algoritma *K Nearest Neighbor* sebanyak tiga kali uji coba, pada pengujian pertama menggunakan 90% pelatihan data dan 10% pengujian data memperoleh akurasi sebesar 91.67%. Pengujian kedua menggunakan 80% pelatihan data dan 20% pengujian data memperoleh akurasi sebesar 92%. Pada pengujian terakhir menggunakan 70% pelatihan data dan 30% pengujian data 90.54%. Dari hasil uji coba untuk kedua algoritma, algoritma *Naive Bayes* memiliki bobot akurasi terbaik sebesar 98.36% dengan 70% pelatihan data dan 30% pengujian data dengan *split validation* dalam melakukan prediksi kepuasan pengguna fitur Tiktok Shop pada aplikasi Tiktok.

## V. SARAN

Penelitian ini dapat dikembangkan lagi untuk menyempurnakan penelitian, saran yang dapat dilakukan adalah sebagai berikut:

1. Dapat menggunakan atau menambahkan atribut lain dalam mengukur kepuasan pengguna.
2. Meningkatkan jumlah dataset yang digunakan, karena jumlah dari dataset dapat mempengaruhi hasil akurasi.
3. Penelitian dapat dikembangkan dengan membandingkan algoritma lain untuk mendapatkan hasil yang bervariasi dan akurasi yang maksimal dalam mengukur kepuasan pengguna.

## REFERENSI

- [1] Abdullah, M.R. *Metode penelitian kuantitatif*. 2015
- [2] Al-Khasawneh, M., et al. "The adoption of TikTok application using TAM model." *International Journal of Data and Network Science* 6, no. 4 (2022).
- [3] Angelina, et al. "Analyzing e-commerce success using DeLone and McLean model." *Journal of Information Systems Engineering and Business Intelligence* 5, no. 2 (2019).
- [4] Annur, Cindy Mutia. *10 Negara dengan Jumlah Pengguna TikTok Terbanyak di Dunia (Januari 2023)*. n.d. <https://databoks.katadata.co.id/datapublish/2023/02/27/indonesia-sabet-posisi-kedua-sebagai-negara-pengguna-tiktok-terbanyak-di-dunia-pada-awal-2023> (tanggal akses 1 Maret, 2023)
- [5] Cahya Dewi A, et al. *Kepuasan Pengguna Layanan Shopee Food Menggunakan Algoritma Naive Bayes* 12, no. 2 (2022).
- [6] Dennis Aprilla C, et al. "Belajar Data Mining dengan RapidMiner." 2013.
- [7] D'Souza, Deborah. "TikTok: What It Is, How It Works, and Why It's Popular." February 2023.
- [8] Ekawati, Ahmad Muharsyah dan Rika Kharlina. "Analisis Pengaruh Kualitas Informasi, Kualitas Sistem dan Kualitas Layanan Terhadap Kepuasan Pengguna Pada Aplikasi Tokopedia Dengan Model Delone And Mclean Di Kota Palembang." *Jurnal Sistem & Teknologi Informasi Komunikasi* 4, no. 2 (2021).
- [9] Fajriyan Nur, et al. "Komparasi Tingkat Akurasi Information Gain dan Gain Ratio pada Metode K-nearest Neighbor." *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2022: Vol. 6 No. 1.
- [10] Grandis, Galuh Fadillah. "Seleksi Fitur Gain Ratio pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh dengan K-Nearest Neighbor." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2021: Vol. 5, No. 8.
- [11] Jirapon Sunkpho, Markus Hofmann. "Analyzing Customer Satisfaction of a Mobile Application." *Thammasat Review* 22, no. 2 (2019).
- [12] Liam O'Brien, et al. "Quality Attributes and Service-Oriented Architectures." (Software Engineering Institute) 2005.
- [13] Lidwina, Andrea. *Penggunaan E-Commerce Indonesia Tertinggi di Dunia*. 2021. <https://databoks.katadata.co.id/datapublish/2021/06/04/penggunaan-e-commerce-indonesia-tertinggi-di-dunia> (tanggal akses Januari 26, 2023).
- [14] M. Syukri Mustafa, et al. "Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyapa Kabupaten Bulukumba." *PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, 2019: Vol. VIII, No. 1
- [15] Muchlis, Lita Sari. "Proses Decision Tree Pada Datamining dengan Algoritma Id3." *Jurnal Sainstek*, 2010: Vol. II No. 1:87-93.
- [16] Populix. *The Social Commerce Landscape in Indonesia*. 2022. <https://info.populix.co/report/the-social-commerce-in-indonesia/> (tanggal akses 26 Januari, 2023).
- [17] Pratomo, Viliane Puspa Negara dan Dudi. "Pengaruh Kualitas Sistem, Kualitas Informasi Dan Kualitas Layanan Her Registrasi Igracias Terhadap Kepuasan Pengguna." *JAF (Journal of Accounting and Finance)* 01, no. 01 (2019)
- [18] Ratnawati, Fajar. "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter." *JURNAL INOVTEK POLBENG* 3, no. 1 (2018)
- [19] Retnawati, H. "Validitas dan Reliabilitas Konstruk Skor Tes Kemampuan Calon Mahasiswa." *Jurnal Ilmu Pendidikan* 23, no. 2 (2017)
- [20] S, T. Yusran. "Penggunaan Metode NAIVE BAYES Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent." *TeKa* 9, no. 2 (2019)
- [21] Sabri, Alia. "Applying DeLone and McLean IS success model on sociotechno knowledge management system." *IJCSI International Journal of Computer Science Issues* 11, no. 6 (2014): 2.
- [22] Seran, Dela Rosari Maria. "Perbandingan Metode K-nearest Neighbour Dan Naive Bayes Classification Untuk Membentuk Sistem Rekomendasi Berita Olahraga Online Berbasis Web Application." *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, 2021
- [23] Siregar, C. Naomi, dkk. "Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar." *Jurnal Teknologi*, 2020: Vol. 3, No. 1.
- [24] Siswanto, Didik. "Customer Satisfaction Level Analysis Of E-commerce Shopee Using Delon And Mclean Methods." *Proceeding 2nd International Conference on Business & Social Sciences (ICOBUS)*, 2022
- [25] Syafina, et al. *Metode Penelitian Akuntansi : Pendekatan Kuantitatif*. 2019.
- [26] T. Hervira, et al. "Analisis Tingkat Kepuasan Pelanggan Terhadap Penggunaan Gmalite dengan Penerapan Algoritma Naive Bayes pada MCDonalds Cibitung." *JSI (Jurnal sistem Informasi) Universitas Suryadarma* 9, no. 2 (2022).
- [27] William H. DeLone, Ephraim R. McLean. "Information Systems Success: The Quest for the Dependent Variable." *Information System Research* 3 (1992): 1.
- [28] William H. DeLone, Ephraim R. McLean. "The DeLone and McLean model of information systems success: a ten-year update." *Journal of management information systems* 19, no. 4 (2003): 9-30.
- [29] Y., Yuningsih. "Metode Delone dan Mclean Dalam Kepuasan Konsumen." *Bina Insani ICT Journal* 6, no. 1 (2019): 55-64.
- [30] Yunita, Devi. "Perbandingan Algoritma K-nearest Neighbor dan Decision Tree untuk Penentuan Risiko Kredit Kepemilikan Mobil ." *JURNAL INFORMATIKA UNIVERSITAS PAMULANG*, 2017: Vol. 2, No. 2.