

# Analisis Sentimen Terhadap TikTok Shop Dengan *K-Nearest Neighbor*, *Decision Tree*, dan *Naive Bayes*

Fieryando<sup>1</sup>, Budhi Kristianto<sup>\*2</sup>

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya  
Wacana, Salatiga 50711, Indonesia

[fieryando32@gmail.com](mailto:fieryando32@gmail.com)<sup>1</sup>, [budhik@uksw.edu](mailto:budhik@uksw.edu)<sup>2</sup>

**Abstract.** *Sentiment Analysis on TikTok Shop using K-Nearest Neighbor, Decision Tree, and Naive Bayes.* In the rapidly developing digital era, many applications help humans, one of which is TikTok. TikTok is an application that displays videos with various categories; TikTok has a marketplace feature, namely TikTok Shop. TikTok Shop is a feature we can use to shop to fulfill our needs. The presence of the TikTok Shop feature gives people opinions regarding this feature, some positive and some negative. With people's opinions about this TikTok Shop feature, research was conducted with three algorithms, namely K-Nearest Neighbor, getting an accuracy result of 94%, Decision Tree with an accuracy of 96%, and Naive Bayes with an accuracy of 98%. This research was conducted to provide excellent and easy-to-understand data related to TikTok Shop user opinions, which might be a basis for the evaluation of TikTok Shop features to be better.

**Keywords:** *TikTok, sentiment analysis, K-Nearest Neighbor, Decision Tree, Naive Bayes*

**Abstrak.** Di era digital yang berkembang dengan cepat, banyak aplikasi yang membantu manusia, salah satunya adalah TikTok. TikTok merupakan aplikasi yang menampilkan video dengan berbagai jenis kategori, TikTok juga terdapat fitur marketplace, yaitu TikTok Shop. TikTok Shop merupakan sebuah fitur yang dapat digunakan untuk berbelanja dalam memenuhi kebutuhan. Hadirnya fitur TikTok Shop membuat orang-orang berpendapat terkait fitur tersebut, ada yang berpendapat positif dan juga negatif. Dengan adanya pendapat dari orang-orang tentang fitur TikTok Shop ini, dilakukan penelitian dengan tiga algoritma, yaitu K-Nearest Neighbor mendapatkan hasil akurasi sebesar 94%, Decision Tree dengan akurasi sebesar 96% dan Naive Bayes dengan akurasi sebesar 98%. Penelitian ini dilakukan untuk menyediakan data yang baik dan mudah dimengerti terkait pendapat pengguna TikTok Shop, sehingga dapat menjadi landasan evaluasi untuk perbaikan fitur TikTok Shop.

**Kata kunci:** *TikTok, analisis sentimen, K-Nearest Neighbor, Decision Tree, Naive Bayes*

## 1. Pendahuluan

Kemajuan teknologi yang cepat menjadikan banyak platform digital yang bermunculan tanpa kita sadari, hal tersebut tentu memiliki pengaruh yang cukup besar terhadap aktivitas manusia. Salah satu platform digital yang telah tersebar luas adalah media sosial, yang merupakan suatu platform di internet di mana pengguna dapat menggambarkan diri mereka sendiri dan berinteraksi secara sosial di dunia maya [1]. Salah satu platform yang memiliki jumlah pengguna yang besar ialah TikTok, yang merupakan sebuah aplikasi berbasis internet yang penggunaannya dapat membuat serta membagikan video singkat [2].

Dina Bhirawa, Direktur Pemasaran TikTok, mengungkapkan bahwa TikTok memiliki 10 juta pengguna aktif setiap bulan di Indonesia. Sementara itu, jumlah pengguna harian mencapai lima juta pengguna. Ternyata, informasi ini kurang dikenal oleh sebagian besar pelaku masyarakat di Indonesia. Penyebabnya karena TikTok masih dianggap sebagai aplikasi baru dan mayoritas penggunaannya adalah remaja dan anak-anak [3]. Seiring berjalannya waktu, TikTok semakin

banyak dikenal di Indonesia baik dari kaum anak-anak, remaja, maupun orang tua sehingga TikTok memunculkan inovasi baru yaitu dengan fitur *e-marketplace* di dalamnya, fitur tersebut dinamakan TikTok Shop.

TikTok Shop ialah sebuah fitur *e-marketplace* baru yang dihadirkan oleh TikTok, fitur ini membantu pengguna untuk dapat berbelanja dalam hal memenuhi kebutuhan sehari-hari maupun jangka panjang tanpa harus keluar rumah. Dengan adanya fitur TikTok Shop ini orang-orang mulai beropini terkait fitur yang diberikan, ada yang beropini positif, negatif, maupun netral, dan bahkan ada juga yang membandingkannya dengan *e-marketplace* yang lain. Dengan adanya opini-opini yang diberikan oleh orang yang telah menggunakan fitur tersebut, maka ingin dilakukan analisis sentimen terhadap opini-opini tersebut guna mengukur seberapa bagus pelayanan fitur yang diberikan.

Dengan beberapa latar belakang tersebut dilakukan penelitian terhadap *e-marketplace* TikTok Shop dengan mengimplementasikan algoritma *K-Nearest Neighbor*, *Decision Tree*, dan *Naive Bayes*. Penelitian dilakukan dengan ketiga metode tersebut karena penelitian sebelumnya mendapatkan hasil yang berbeda, maka ingin diketahui seberapa besar pengaruh ketiga metode tersebut terhadap TikTok Shop. Penelitian ini dilakukan dengan mengumpulkan data melalui Twitter. Twitter merupakan sebuah platform media sosial yang menjadi tempat menyampaikan opini atau pendapat pribadi yang digunakan oleh masyarakat dari berbagai negara. Penggunaan Twitter di Indonesia berkembang dengan cepat dan menempatkan Indonesia pada peringkat kelima di dunia [4]. *Tools* yang dimanfaatkan dalam penelitian ini adalah Rapidminer, untuk mengolah data Twitter yang telah didapatkan dengan menghitung akurasi yang didapatkan dari metode yang ingin digunakan. Selain itu, tujuan penelitian ini agar bisa memahami informasi yang didapatkan dengan lebih mudah terkait opini tentang TikTok Shop yang dapat menjadi evaluasi untuk fitur TikTok Shop agar dapat ditingkatkan lebih lanjut.

## 2. Tinjauan Pustaka

### 2.1. Analisis Sentimen

Sebuah hal yang dilakukan untuk membuat opini yang diberikan dari pengguna aplikasi, sentimen analisis juga mewujudkan salah satu proses yang digunakan untuk mengekstrak, membuat data tekstual dan menangkap data. Ada tiga jenis opini yang sering ditemui yaitu netral, positif, dan negatif, dengan opini tersebut perusahaan dapat mempertimbangkan atau menganalisis kembali produk yang mereka buat [5].

### 2.2. E- Marketplace

*E-marketplace* adalah sarana yang digunakan agar memasarkan layanan melalui internet. Dalam konteks ini, “*place*” mengacu pada tempat atau platform yang memungkinkan pemasaran produk atau layanan melalui internet [6].

### 2.3. Analisis Penelitian Terdahulu

Penelitian mengenai analisis sentimen telah dilakukan oleh beberapa peneliti terdahulu dengan menggunakan studi kasus yang berbeda, yang dapat dilihat pada Tabel 1.

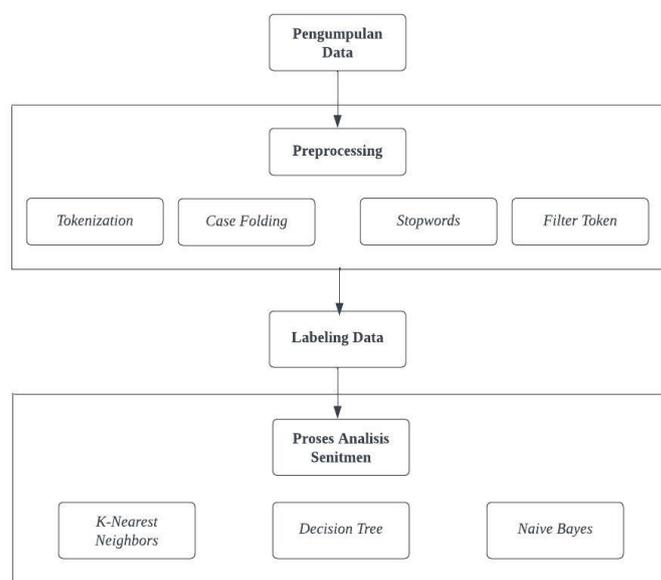
Penelitian yang dilakukan oleh Saputra dkk., bertujuan untuk mengukur seberapa baik pelayanan yang telah diberikan oleh *e-marketplace* Tokopedia. Dalam proses analisis yang dilakukan, peneliti mengambil pendapat atau opini orang-orang melalui Twitter menggunakan *data mining*, dan menganalisisnya menggunakan metode *Decision Tree*, K-NN, dan *Naive Bayes*. Dengan menggunakan ketiga metode tersebut, peneliti mendapatkan hasil yang paling akurat dari *Decision Tree* yaitu sebesar 82% [7]. Menurut penelitian yang dilakukan oleh Limbong tentang analisis sentimen terhadap *e-marketplace* Shopee dengan mengimplementasikan algoritma *Naive Bayes* dan K-NN, serta pengumpulan data melalui Google Playstore, peneliti mendapatkan hasil di mana K-NN mendapatkan hasil akurasi yang lebih akurat yaitu sebesar 92% [8].

Penelitian lainnya dilakukan oleh Bayhaqy dkk. Penelitian tersebut membahas tentang analisis sentimen terkait Tokopedia dan Bukalapak berdasarkan data dari Twitter dengan mengimplementasikan algoritma *Decision Tree*, K-NN, dan *Naive Bayes*. Para peneliti

mendapatkan hasil paling akurat dengan algoritma *Naive Bayes* yaitu mendapatkan hasil akurasi 80%. Pada penelitian ini disebutkan bahwa pengklasifikasian *Naive Bayes* merupakan metode yang paling akurat dengan *dataset* media sosial [9]. Namun ada juga penelitian lainnya yang dilakukan oleh Firdaus dkk. Penelitian ini membahas mengenai analisis sentimen mengenai Tokopedia dengan menggunakan ulasan dari Google Playstore dengan metode *Naive Bayes* dan K-NN. Peneliti mendapatkan hasil bahwa K-NN mendapatkan hasil akurasi yang lebih akurat dibandingkan dengan *Naive Bayes* yaitu sebesar 86% [10]. Melihat beberapa analisis penelitian terdahulu mengenai analisis sentimen dengan *dataset* dan algoritma yang berbeda serta mendapatkan hasil yang berbeda pula, penulis tertarik untuk melakukan analisis sentimen terhadap TikTok Shop dengan menggunakan algoritma *K-Nearest Neighbor*, *Decision Tree*, dan *Naive Bayes*. Penelitian ini dibuat dengan ide sendiri dan beberapa penelitian terdahulu sebagai acuan untuk menganalisis sentimen terhadap objek yang berbeda yaitu TikTok Shop.

**Tabel 1. Analisis Penelitian Terdahulu**

No	Judul Penelitian	Metode	Perbedaan
1	Analisis Sentimen Pengguna <i>Marketplace</i> Bukalapak dan Tokopedia di Twitter Menggunakan <i>Machine Learning</i> [7]	<i>Decision Tree</i> , <i>K-Nearest Neighbor</i> , dan <i>Naive Bayes</i>	<i>Dataset</i> yang digunakan dalam penelitian berbeda
2	Analisis Klasifikasi Sentimen Ulasan pada <i>E-Commerce</i> Shopee Berbasis <i>Word Cloud</i> dengan Metode <i>Naive Bayes</i> dan <i>K-Nearest Neighbor</i> [8]	<i>Naive Bayes</i> dan <i>K-Nearest Neighbor</i>	<i>Dataset</i> dan algoritma yang digunakan dalam penelitian berbeda
3	<i>Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naive Bayes</i> [9]	<i>Decision Tree</i> , <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	<i>Dataset</i> yang digunakan dalam penelitian berbeda
4	Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma <i>Naive Bayes Classifier</i> dan <i>K-Nearest Neighbor</i> [10]	<i>Naive Bayes</i> dan <i>K-Nearest Neighbor</i>	<i>Dataset</i> dan Algoritma yang digunakan dalam penelitian berbeda



**Gambar 1. Metodologi Penelitian**

### 3. Metode Penelitian

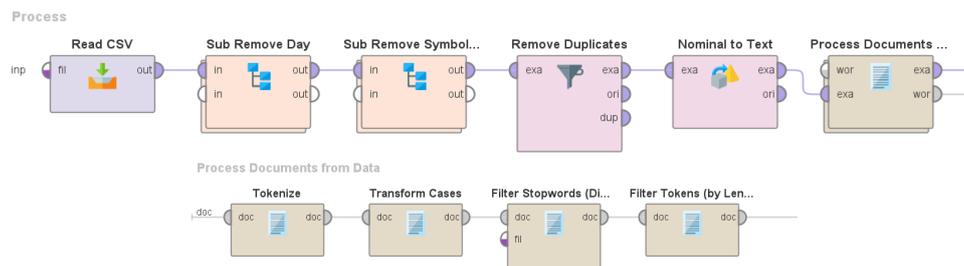
Proses penelitian yang dilakukan dijelaskan dalam Gambar 1. Tahap pertama metodologi yaitu dilakukan pengumpulan data, tahap ini dilakukan dengan memanfaatkan media sosial Twitter dan Google Colab. Setelah didapatkan data, tahap selanjutnya yaitu dilakukan *preprocessing* agar data yang didapatkan dapat diolah tanpa adanya gangguan. Selanjutnya yaitu *labeling data*, *labeling data* ini yaitu proses pengelompokan data yang didapatkan, agar nantinya dapat mengetahui hasil sentimen terhadap sebuah ulasan. Tahap terakhir yaitu dilakukan analisis sentimen dengan menerapkan algoritma K-NN, *Decision Tree*, dan *Naive Bayes*.

### 3.1. Pengumpulan Data

Informasi berupa pendapat langsung masyarakat mengenai TikTok Shop diperoleh dari platform media sosial Twitter. Data dikumpulkan dengan *web crawler* menggunakan bahasa pemrograman Python dan Twitter dengan *auth token* dari setiap akun agar dapat mengumpulkan data dengan limit 500 dalam setiap pengumpulan datanya dari setiap akun. *Auth token* dapat diperoleh dari setiap masing-masing akun Twitter.

### 3.2. Preprocessing

*Preprocessing* adalah langkah pengolahan data teks yang melibatkan penghapusan informasi yang tidak diperlukan agar memungkinkan analisis sentimen selanjutnya [11]. Tahap *preprocessing* ini sangat diperlukan agar dapat menghasilkan data yang berkualitas dan mendapatkan keputusan yang ideal. Tahap *preprocessing* digambarkan pada Gambar 2.



**Gambar 2. Tahap *Preprocessing***

Terdapat empat tahap *preprocessing*, yaitu *tokenization*, *case folding*, *stopwords*, dan *filter tokens*. *Tokenization* merupakan proses yang melibatkan pembagian atau pemecahan kalimat menjadi bagian-bagian yang disebut sebagai token. Apabila terdapat tanda baca yang tidak relevan, maka akan dihapus [12]. *Case folding* adalah tahap mengganti huruf yang terdapat pada kalimat menjadi huruf kecil. *Stopwords* adalah tahap yang akan menghapus kata yang tidak penting dalam sebuah kalimat. *Filter tokens* merupakan proses untuk menghapus kata dengan ketentuan panjang dari kalimat tersebut. Penelitian ini menggunakan ketentuan minimal empat karakter dan maksimal 25 karakter agar dapat menghilangkan kalimat yang tidak informatif, sedangkan maksimal 25 agar kalimat tersebut tidak terlalu panjang, hanya mengambil inti dari kalimat saja.

### 3.3. Labelling Data

Proses *labelling data* dilakukan untuk memisahkan data menjadi beberapa sentimen terhadap suatu kalimat. Penelitian ini terdapat tiga sentimen, berupa positif, negatif, dan netral. Hal ini dilakukan dengan tujuan nantinya data tersebut dapat dibagi menjadi dua bagian, yaitu *data training* dan *data testing*. *Data training* adalah data yang dimanfaatkan untuk melatih sistem agar mampu mengidentifikasi pola yang diinginkan, sedangkan *data testing* adalah data yang digunakan untuk mencoba kinerja dari pelatihan yang telah dilakukan [13]. Dalam proses *labelling data* terdapat proporsi untuk melakukan *data training* dan *data testing*, *data training* memiliki proporsi lebih besar dibandingkan *data testing*, namun pada umumnya yang digunakan adalah 70:30, 80:20, atau 90:10 [14]. Penelitian ini menggunakan 80:20; 80 adalah *data testing* dan 20 adalah *data training*.

### 3.4. Proses Analisis Sentimen

Dalam penelitian ini, ketiga metode dibandingkan karena ingin mengetahui seberapa besar pengaruh setiap metode terhadap ulasan TikTok Shop. Ditemukan juga hasil yang tidak sama berdasarkan penelitian-penelitian yang sudah ada, ketidaksamaan hasil tersebut digunakan sebagai acuan. Ketiga algoritma yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor*, *Decision Tree*, dan *Naive Bayes*. *K-Nearest Neighbor* adalah algoritma klasifikasi berdasarkan

data pembelajaran dengan mengelompokkan objek berdasarkan atribut, beserta sampel pelatihan yang berjarak paling dekat dengan objek tersebut [15]. *Decision Tree* adalah metode untuk mengklasifikasikan data. Algoritma *Decision Tree* dirancang untuk menciptakan model prediksi yang sangat spesifik dalam bentuk aturan yang dapat dengan mudah dijalankan [16]. *Naive Bayes* adalah metode klasifikasi yang menggunakan probabilitas dan statistik. Prinsip dasarnya adalah mengukur probabilitas yang akan datang berdasarkan pengalaman lampau, sehingga sering disebut sebagai Teorema Bayes [17].

## 4. Hasil dan Diskusi

### 4.1. Pengumpulan Data

Informasi yang berhasil dikumpulkan sebanyak 664 dari 6.826 data. Data tersebut diambil dari Twitter dengan menggunakan *web crawler* bahasa pemrograman Python. Jangka waktu yang digunakan dalam pengumpulan data tersebut ialah dimulai dari tanggal 1 Januari 2023 sampai dengan 31 Maret 2023. Dilakukan analisis pemilihan data pada 6.826 data ini dan mendapatkan hasil data sebanyak 664 yang dapat digunakan dalam penelitian ini karena data yang diambil hanya berupa ulasan terkait TikTok Shop. Banyak sebaran data dari 6.826 yang hanya merupakan promosi dan rekomendasi barang yang tidak dibutuhkan dalam penelitian ini.

### 4.2. Preprocessing

Berikut tahap-tahap yang dilakukan dalam *preprocessing* data guna mendapatkan hasil data yang berkualitas dan mendapatkan keputusan yang ideal.

#### 4.2.1. Tokenization

Pada proses ini dilakukan pembagian kata dari kalimat yang telah didapatkan, berikut contoh dari *tokenization* pada Tabel 2. Pada Tabel 2 menunjukkan *Total Occurences* sebanyak 673, dan *Document Occurences* sebanyak 637, yang artinya sebanyak 673 kata TikTok muncul dalam 637 kalimat yang terdapat dalam dokumen.

**Tabel 2. Contoh Tokenization**

<i>Word</i>	<i>Attribute Name</i>	<i>Total Occurences</i>	<i>Document Occurences</i>
tiktok	tiktok	673	637
shop	shop	661	652
murah	murah	121	108
belanja	belanja	84	79
beli	beli	75	70
racun	racun	72	72
banget	banget	70	68

#### 4.2.2. Case Folding

*Case folding* mengubah semua teks yang hurufnya tidak beraturan, di mana ada huruf besar, dan juga huruf kecil dalam suatu teks dalam kalimat. Tahap ini membuat semua huruf tersebut menjadi huruf kecil agar nantinya mudah untuk diproses. Berikut contoh dari *case folding* dapat dilihat pada Tabel 3.

**Tabel 3. Sebelum dan Setelah Case Folding**

<i>Sebelum Case Folding</i>	<i>Setelah Case Folding</i>
THANK YOUU TIKTOK SHOP HUEHUE	thank youu tiktok shop huehue
gua ga berani belanja lewat tiktok shop dah	gue ga berani belanja lewat tiktok shop dah
Bahaya emang kalo emak main tiktok shop	bahaya emang kalo emak main tiktok shop
Tiktok shop lama bgt anjer pengirimannya	tiktok shop lama bgt anjer pengirimannya
Tiktok shop ternyata cair uang nya lama ya Seminggu dr pesanan sampai	tiktok shop ternyata cair uang nya lama ya seminggu dr pesanan sampai
Nontonin jualan di live Tiktok shop kok seru bgt ya	nontonin jualan di live tiktok shop kok seru bgt ya

#### 4.2.3. Stopwords

*Stopwords* adalah proses untuk mengurangi banyaknya kata yang tidak berguna dalam proses nantinya. Berikut contoh dari *stopwords* dapat dilihat pada Tabel 4. Tabel 4 merupakan

contoh proses *preprocessing stopwords* dengan menghilangkan beberapa kata yang tidak diperlukan dalam proses analisis. Contohnya pada baris ke tujuh kata “ada” dihilangkan setelah *stopwords*. Tabel 5 menunjukkan bahwa sebanyak 1.684 data sebelum dilakukan *stopwords* dan menghasilkan sebanyak 1.463 data yang tersisa setelah dilakukan *stopwords*.

**Tabel 4. Sebelum dan Setelah *Stopwords***

Sebelum <i>Stopwords</i>				Setelah <i>Stopwords</i>			
<i>Word</i>	<i>Attribute Name</i>	<i>Total Occurrences</i>	<i>Document Occurrences</i>	<i>Word</i>	<i>Attribute Name</i>	<i>Total Occurrences</i>	<i>Document Occurrences</i>
aaa	aaa	1	1	aaa	aaa	1	1
aamiin	aamiin	1	1	aamiin	aamiin	1	1
abal	abal	2	1	abal	abal	2	1
abis	abis	4	4	abis	abis	4	4
account	account	1	1	account	account	1	1
aci	aci	1	1	aci	aci	1	1
ada	ada	30	26	adiktif	adiktif	1	1
adalah	adalah	5	5	admin	admin	1	1
adiktif	adiktif	1	1				
admin	admin	1	1				

**Tabel 5. Hasil Sebelum dan Setelah *Stopwords***

Sebelum <i>Stopwords</i>	Setelah <i>Stopwords</i>
<i>ExampleSet (664 examples, 0 special attributes, 1.648 regular attributes)</i>	<i>ExampleSet (664 examples, 0 special attributes, 1.648 regular attributes)</i>

#### 4.2.4. Filter Tokens

Penelitian ini menggunakan *filter tokens* dengan jumlah kalimat minimal empat kata dan maksimal sebanyak 25 kata. Tabel 6 menunjukkan contoh dari *filter tokens*.

**Tabel 6. Sebelum dan Setelah *Filter Tokens***

Sebelum <i>Filter Tokens</i>				Setelah <i>Filter Tokens</i>			
<i>Word</i>	<i>Attribute Name</i>	<i>Total Occurrences</i>	<i>Document Occurrences</i>	<i>Word</i>	<i>Attribute Name</i>	<i>Total Occurrences</i>	<i>Document Occurrences</i>
aaa	aaa	1	1	aamiin	aamiin	1	1
aamiin	aamiin	1	1	abal	abal	2	1
abal	abal	2	1	abis	abis	4	4
abis	abis	4	4	account	account	1	1
account	account	1	1	adiktif	adiktif	1	1
aci	aci	1	1	admin	admin	1	1
adiktif	adiktif	1	1				
admin	admin	1	1				

#### 4.3 Labelling Data

*Labelling data* dilakukan dengan cara membuat *data training* dan *data testing* sebesar 20:80. Proses dilakukan secara manual untuk 20% data, selanjutnya sisa data sebesar 80% dilakukan menggunakan *machine learning*. Contoh *labelling data* dapat dilihat pada Tabel 7.

**Tabel 7. Labeling Data**

<i>Row No.</i>	<i>sentimen</i>	<i>predict (sentimen)</i>	<i>confidence (positif)</i>	<i>confidence (netral)</i>	<i>confidence (negatif)</i>	<i>text</i>
1	?	positif	1	0	0	tabungan an...
2	?	positif	1	0	0	gila tiktok sho...
3	?	netral	0	0	1	tiktok live sho...
4	?	netral	0	0	1	ketagihan no...
5	?	negatif	0	1	0	ketagihan pa...

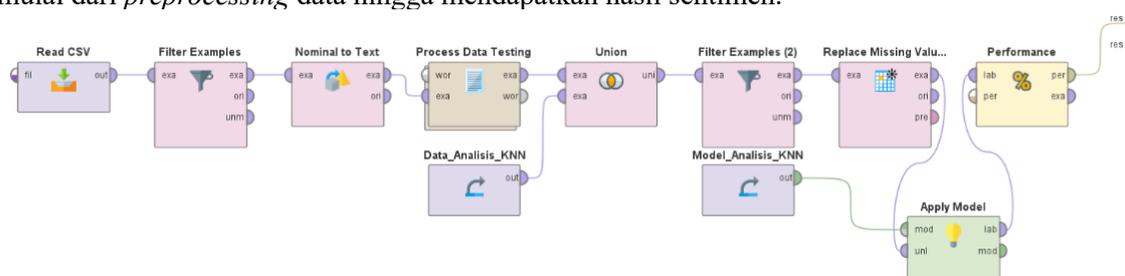
#### 4.4 Hasil Analisis Sentimen

Setelah melalui tahap di atas, maka didapatkanlah hasil dari setiap metode yang digunakan, dengan skema yang telah dibuat menggunakan Rapidminer. Sebelum memasuki hasil

analisis, perlu diketahui dalam setiap hasil dari analisis sentimen terdapat pred.positif, pred.negatif, dan pred.netral. Pred.positif adalah banyaknya prediksi data yang sentimennya sama dengan *true* positifnya. Pred.negatif adalah banyaknya prediksi data yang sentimennya sama dengan *true* negatifnya. Pred.netral adalah banyaknya prediksi data yang sentimennya sama dengan *true* netralnya. Sebagai contoh, pada Tabel 8 terdapat 590 prediksi positif yang sama dengan *true* positif, terdapat 48 prediksi negatif yang sama dengan *true* negatif, dan terdapat 2 prediksi netral yang sama dengan *true* netral.

#### 4.4.1. K-Nearest Neighbors (KNN)

Skema *K-Nearest Neighbors* (KNN) yang digunakan seperti yang ditunjukkan pada Gambar 3. Pada tahap awal terdapat beberapa proses yang telah dijelaskan sebelumnya di atas, mulai dari *preprocessing* data hingga mendapatkan hasil sentimen.



Gambar 3. Skema *K-Nearest Neighbors*

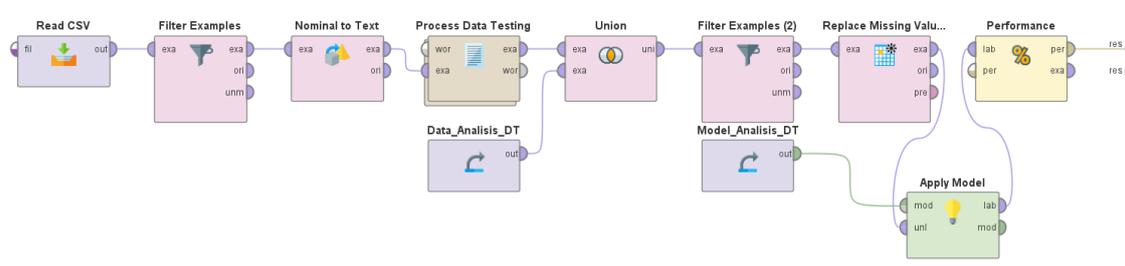
Dengan menggunakan skema pada Tabel 8, *data training* dan model yang digunakan pada metode K-NN mendapatkan hasil akurasi sebesar 96,08%. Hasil akurasi *class precision* pred.positif sebesar 96,56%, pred.negatif 90,20%, dan pred.netral 100%, sedangkan untuk *class recall* memiliki *true* positif 99,33%, *true* negatif 76,67%, dan *true* netral 20,00%.

Tabel 8. Hasil Analisis Sentimen *K-Nearest Neighbor*

<i>Accuracy: 96,08%</i>	<i>True Positif</i>	<i>True Negatif</i>	<i>True Netral</i>	<i>Class Precision</i>
<b>Pred. Positif</b>	590	14	7	96,56%
<b>Pred. Negatif</b>	4	46	1	90,20%
<b>Pred. Netral</b>	0	0	2	100,00%
<b>Class Recall</b>	99,33%	76,67%	20,00%	

#### 4.4.2. Decision Tree

Skema *decision tree* ditunjukkan pada Gambar 4. Tahapan yang dilakukan serupa dengan metode KNN, mulai dari *preprocessing* hingga *labeling data*.



Gambar 4. Skema *Decision Tree*

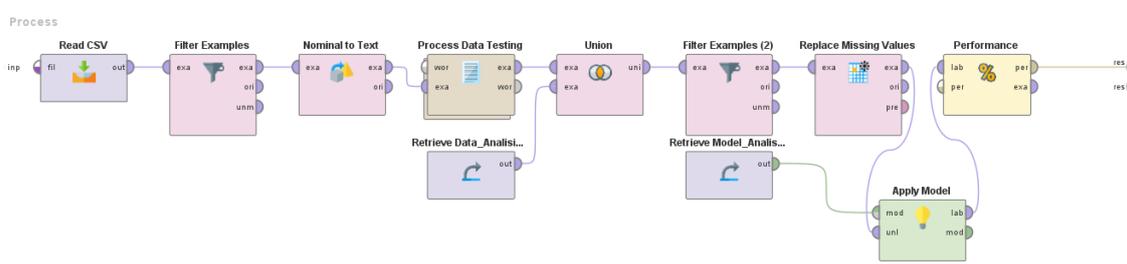
Dengan menggunakan skema pada Gambar 4, *data training* dan model yang digunakan mendapatkan hasil akurasi sebesar 94,13%. Tabel 9 menunjukkan hasil dari analisis sentimen menggunakan *decision tree*; *class precision* pred.positif sebesar 94,13%, pred.negatif 0,00%, dan pred.netral 0,00% sedangkan untuk *class recall* memiliki nilai *true* positif 100,00%, *true* negatif 0,00%, dan *true* netral 0,00%.

**Tabel 9. Hasil Analisis Sentimen *Decision Tree***

<b>Accuracy: 94,13%</b>	<b>True Positif</b>	<b>True Negatif</b>	<b>True Netral</b>	<b>Class Precision</b>
<b>Pred. Positif</b>	625	31	8	94,13%
<b>Pred. Negatif</b>	0	0	0	0.00%
<b>Pred. Netral</b>	0	0	0	0.00%
<b>Class Recall</b>	100,00%	0,00%	0,00%	

#### 4.4.3. *Naive Bayes*

Skema pada metode *Naive Bayes* ditunjukkan pada Gambar 5. Dengan menggunakan skema pada Gambar 5, *data training* dan model yang digunakan mendapatkan hasil akurasi sebesar 98,34%, seperti ditampilkan pada Tabel 10. Nilai akurasi untuk *class precision* pred.positif adalah sebesar 100,00%, pred.negatif sebesar 95,04%, dan pred.netral sebesar 96,43%, sedangkan untuk *class recall* memiliki nilai *true positif* 97,39%, *true negatif* 100,00%, dan *true netral* 100,00%.

**Gambar 5. Skema *Naive Bayes*****Tabel 10. Hasil Analisis Sentimen *Naive Bayes***

<b>Accuracy: 98.34%</b>	<b>True Positif</b>	<b>True Negatif</b>	<b>True Netral</b>	<b>Class Precision</b>
<b>Pred. Positif</b>	411	0	0	100.00%
<b>Pred. Negatif</b>	7	134	0	95.04%
<b>Pred. Netral</b>	4	0	108	96.43%
<b>Class Recall</b>	97.39%	100.00%	100.00%	

## 5. Kesimpulan dan Saran

Penelitian ini menggunakan *tools* Rapidminer versi 10.2.0 dan metode *K-NN*, *Decision Tree*, serta *Naive Bayes*. Data yang digunakan sebagai *dataset* didapatkan dari media sosial Twitter dengan menggunakan *auth token*. Berdasarkan penelitian yang telah dilakukan terdapat sebanyak 664 data mengenai TikTok Shop dapat disimpulkan bahwa metode *Naive Bayes* mendapatkan akurasi terbaik, yaitu sebesar 98,34%. Metode *K-Nearest Neighbors* hanya memiliki akurasi sebesar 96,08% dan metode *decision tree* sebesar 94,13%. Penelitian berikutnya dapat dikembangkan dengan menggunakan jumlah informasi data yang lebih banyak agar memiliki hasil yang lebih akurat, dan dapat menggunakan metode lain sebagai pembandingan.

## Referensi

- [1] A. Setiadi, "Pemanfaatan media sosial untuk efektivitas komunikasi," *Cakrawala: Jurnal Humaniora Bina Sarana Informatika*, vol. 16, no. 2, 2016, doi: 10.31294/jc.v16i2.1283.
- [2] P. Ghosh and A. Ghosh, "An unusual case of video app addiction presenting as withdrawal psychosis," *Addiction*, vol. 12, no. 1, pp. 40455-40457, 2021, doi: 10.24327/ijrsr.2021.1201.5700.
- [3] A. M. Damar, "Upaya TikTok Edukasi Pengguna Bagian Konten Positif", *Liputan 6*, Jan 23, 2019. [Online], Tersedia : <https://www.liputan6.com/tekno/read/3877620/upaya-tik-tok-edukasi-pengguna-bagikan-konten-positif>. [Diakses pada Sept. 28, 2023].
- [4] I. Utami dan M. Marzuki, "Analisis sistem informasi banjir berbasis media Twitter," *Jurnal Fisika Unand*, vol. 9, no. 1, pp. 67-72, 2020, doi: 10.25077/jfu.9.1.67-72.2020.

- [5] N. T. Romadloni, I. Santoso, dan S. Budilaksono, "Perbandingan Metode Naive Bayes, K-NN Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1-9, 2019.
- [6] W. Y. Sadgotra dan E. H. Saputra, "Perancangan online marketplace untuk usaha kecil dan menengah (UKM) di Kabupaten Purworejo," *Data Manajemen Dan Teknologi Informasi (DASI)*, vol. 14, no. 4, pp. 54-58, 2014.
- [7] I. Saputra, R. S. A. Pambudi, H. E. Darono, F. Amsury, M. R. Fahdia, B. Ramadhan, dan A. Ardiansyah, "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning," *Faktor Exacta*, vol. 13, no. 4, pp. 200-207, 2021, doi: 10.30998/faktorexacta.v13i4.7074.
- [8] J. J. A. Limbong, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor," Doctoral dissertation, Universitas Kristen Satya Wacana, Salatiga, Indonesia, 2022.
- [9] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, dan E. R. Kaburuan, "Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes," in *2018 International Conference on Orange Technologies (ICOT)*, Oct. 2018, pp. 1-6.
- [10] M. F. El Firdaus, N. Nurfaizah, dan S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, pp. 1329-1336, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [11] A. S. Widagdo, B. S. WA, dan A. Nasiri, "Analisis tingkat kepopuleran e-commerce di Indonesia berdasarkan sentimen sosial media menggunakan metode naïve bayes," *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, vol. 6, no. 1, pp. 1-5, 2020, doi: 10.46808/informa.v6i1.159.
- [12] R. Rasenda, H. Lubis, dan R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, pp. 369-376, 2020, doi: 10.30865/mib.v4i2.2051.
- [13] N. M. S. Hadna, P. I. Santosa, dan W. W. Winarno, "Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter," dalam *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp. 57-64, 2016.
- [14] P. A. Jusia, F. M. Irfan, dan K. Kurniabudi, "Clustering data untuk rekomendasi penentuan jurusan perguruan tinggi menggunakan metode K-Means," *IKRA-ITH Informatika: Jurnal Komputer dan Informatika*, vol. 3, no. 3, pp. 75-84, 2019.
- [15] D. Noviana, Y. Susanti, dan I. Susanto, "Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma C4.5," dalam *Seminar & Conference Proceedings of UMT*, Jun. 2019.
- [16] N. Noviandi, "Implementasi algoritma decision tree C4.5 untuk prediksi penyakit diabetes," *Indonesian of Health Information Management Journal (INOHIM)*, vol. 6, no. 1, pp. 1-5, 2018.
- [17] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, dan I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 11, no. 3, pp. 205-215, 2022.