



## Sentimen Analisis Terhadap Fitur Tiktok Shop Menggunakan Naïve Bayes dan K-Nearest Neighbor

Sandi Saputra Hasibuan\*, Angraini, Eki Saputra, Megawati

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>11950313363@students.uin-suska.ac.id, <sup>2</sup>angraini@uin-suska.ac.id, <sup>3</sup>eki.saputra@uin-suska.ac.id, <sup>4</sup>megawati@uin-suska.ac.id

Email Penulis Korespondensi: 11950313363@students.uin-suska.ac.id

**Abstrak**—Salah satu fitur yang diperkenalkan oleh platform TikTok adalah TikTok Shop, di mana penjual dan pembeli dapat berinteraksi dengan mudah melalui fitur siaran langsung dan mempromosikan produk di dalam aplikasi TikTok tanpa perlu beralih ke aplikasi lain, serta menawarkan produk dengan harga yang sangat murah. Namun, hal ini menimbulkan ketidakpuasan dari pedagang kecil yang merasa kesulitan bersaing. Opini tentang fitur TikTok Shop menciptakan berbagai tanggapan dari masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap fitur TikTok Shop dengan menggunakan Twitter sebagai sumber data. Menerapkan algoritma Naive Bayes Classifier (NBC) dan K-Nearest Neighbor (KNN) dan Pembagian data menggunakan 10-Fold Cross Validation. Pelabelan dilakukan dengan metode clustering menggunakan algoritma K-Means, dibagi menjadi tiga kategori positif, negatif, dan netral. Penambahan teknik analisis data kualitatif dengan metode thematic analysis pada penelitian ini bertujuan untuk menemukan pola atau tema dalam data. Dari hasil pelabelan menunjukkan bahwa 75% dari total data mengekspresikan sentimen negatif terhadap fitur TikTok Shop. Dan hasil thematic analysis ditemukan Tema utama yaitu "Regulasi yang tidak sesuai" mencakup 32% data, dengan jumlah 754. Penelitian ini menyimpulkan bahwa metode KNN lebih unggul daripada NBC, dengan akurasi, presisi, dan recall yang lebih baik. Hal ini sejalan dengan penelitian yang dilakukan sebelumnya bahwa KNN lebih unggul dari pada naïve bayes, tetapi penelitian yang lain menunjukkan hal sebaliknya dimana naïve bayes lebih unggul dibandingkan dengan KNN. Penelitian lebih lanjut dapat dilakukan untuk meningkatkan kinerja kedua algoritma ini dalam sentimen analisis, misalnya dengan menggunakan metode preprocessing yang lebih canggih, fitur ekstraksi yang lebih representatif, atau teknik optimisasi yang lebih efisien.

**Kata Kunci:** Analisis Sentimen; Analisis Data Kualitatif; K-Nearest Neighbor; Naive Bayes Classifier; TikTok Shop.

**Abstract**—One of the features introduced by the TikTok platform is TikTok Shop, where sellers and buyers can interact easily through the live broadcast feature and promote products within the TikTok application without having to switch to other applications, as well as offering products at very cheap prices. However, this causes dissatisfaction from small traders who find it difficult to compete. Opinions about the TikTok Shop feature have created various responses from the public. This research aims to analyze public sentiment towards the TikTok Shop feature using Twitter as a data source. Applying the Naive Bayes Classifier (NBC) and K-Nearest Neighbor (KNN) algorithms and dividing data using 10-Fold Cross Validation. Labeling was carried out using the clustering method using the K-Means algorithm, divided into three categories positive, negative and neutral. The addition of qualitative data analysis techniques with thematic analysis method in this research aims to find patterns or themes in the data. The labeling results show that 75% of the total data expresses negative sentiment towards the TikTok Shop feature. And the results of the thematic analysis found that the main theme, namely "Inappropriate regulations", covers 32% of the data, with a total of 754. This research concludes that the KNN method is superior to NBC, with better accuracy, precision and recall. This is in line with previous research that KNN is superior to naïve Bayes, but other research shows the opposite where naïve Bayes is superior to KNN. Further research can be carried out to improve the performance of these two algorithms in sentiment analysis, for example by using more sophisticated preprocessing methods, more representative feature extraction, or more efficient optimization techniques.

**Keywords:** Sentiment Analysis; K-Nearest Neighbor; Naive Bayes Classifier; TikTok Shop.

### 1. PENDAHULUAN

Kemajuan teknologi telah mengubah pola hidup masyarakat menjadi lebih praktis, instan, dan cepat[1]. Belanja online adalah kegiatan transaksi yang dilakukan melalui platform perbelanjaan online atau jejaring sosial yang menawarkan berbagai barang dan jasa. Banyak orang telah mengadopsi kebiasaan berbelanja online karena kenyamanan yang diberikannya[2]. Tiktok shop merupakan salah satu fitur yang disediakan oleh tiktok, tiktok sendiri didirikan oleh Luyu Yang dan Alex Zhu di Shanghai, China pada tahun 2014 dan diperkenalkan oleh ByteDance pada tahun 2017[3]. Popularitas TikTok meningkat seiring waktu karena sejak awal, TikTok telah menjadi platform yang menghadirkan pengalaman baru dengan menggabungkan interaksi sosial media dan kemampuan bertransaksi online[4]. Fitur ini memungkinkan pengguna melakukan pembelian produk secara langsung melalui platform ini. Pemanfaatan fitur TikTok Shop meningkatkan ketertarikan konsumen karena kemudahan dalam proses pembelian, sistem pembayaran yang simpel, dan waktu transaksi yang lebih efisien[5]. Selain itu, di dalam fitur TikTok Shop sendiri penjual dan pelanggan bisa saling menghibur dengan melakukan siaran langsung dan membuat konten video, ini memungkinkan pedagang, merek, dan pembuat konten untuk mempromosikan dan menjual barang secara langsung melalui TikTok Live yang tentunya menjadi peluang besar untuk menjalankan bisnis[4].

Beragamnya opini yang disampaikan oleh masyarakat melalui media sosial, khususnya melalui postingan atau tweet di Twitter, mengenai TikTok Shop menunjukkan kompleksitas pandangan yang ada. Dalam konteks



ini, permasalahan penelitian muncul sehubungan dengan pemahaman mendalam terhadap pendapat masyarakat Indonesia terhadap fitur TikTok Shop. Oleh karena itu, penelitian ini memfokuskan pada eksplorasi dan analisis sentimen sebagai teknik utama untuk memahami opini yang tersebar di masyarakat terkait dengan penggunaan fitur TikTok Shop di Indonesia. Dengan menggunakan teknik analisis sentimen, diharapkan dapat terungkap lebih jelas pola, preferensi, dan persepsi masyarakat terhadap fitur belanja online yang disajikan oleh TikTok.

Analisis sentimen adalah pendekatan analitis yang digunakan untuk menganalisis suatu teks. Tujuan dari analisis sentimen ini adalah untuk mengidentifikasi sejauh mana subjektivitas dari pendapat, hasil review, atau tweet[6]. Analisis sentimen, juga dikenal sebagai penambahan opini, mengamati ekspresi perasaan individu terhadap topik tertentu[7]. Pengambilan sentimen dari tweet ini bermanfaat bagi organisasi dan asosiasi. Sebagai contoh, dapat efektif digunakan sebagai sub-modul pada sistem rekomendasi dan lainnya[8]. Twitter sendiri dipilih untuk pengambilan data karena Twitter adalah platform media sosial yang efektif untuk mendeteksi sentimen publik terhadap suatu topik[9].

Penelitian sebelumnya tentang sentimen analisis terhadap data tweet pengguna tentang layanan transportasi online. Penelitian tersebut menggunakan data tweet sebanyak 565, dan menggunakan metode preprocessing seperti cleaning, tokenizing, filtering dan stemming. Pada penelitian tersebut disimpulkan bahwa algoritma KNN memberikan akurasi yang lebih baik dalam klasifikasi sentimen dibandingkan NBC[10]. Penelitian lainnya mengenai kepuasan pelanggan pembayaran digital di Indonesia, penelitian tersebut menemukan bahwa KNN memiliki akurasi yang lebih baik daripada Naive Bayes[11]. Selanjutnya penelitian terhadap kepuasan layanan pembayaran digital di Indonesia, menyimpulkan klasifikasi menggunakan KNN memiliki akurasi yang lebih tinggi yaitu 90% dibandingkan naive bayes[12]. Penelitian lainnya yaitu penelitian mengenai Perbandingan Klasifikasi Karakter Individu pada Twitter, penelitian tersebut menyimpulkan bahwa algoritma NBC terbukti lebih unggul dari pada algoritma KNN dalam mengklasifikasikan teks dengan pembobotan TF-IDF[13].

Penelitian ini menggunakan algoritma Naïve Bayes Classifier (NBC) dan K-Nearest Neighbor (KNN). Naive bayes merupakan salah satu metode klasifikasi terbanyak yang dapat digunakan untuk analisis sentimen. metode ini menghitung probabilitas pada setiap faktor, lalu memilih hasil dengan kemungkinan nilai tertinggi. ada banyak penerapan algoritma ini seperti klasifikasi teks, pemfilteran spam, analisis sentimen dan lain sebagainya. NBC dapat membuat prediksi secara realtime. KNN merupakan salah satu metode machine learning yang dapat mengklasifikasikan objek berdasarkan pembelajaran data terdekat dengan objeknya. metode ini sangat sederhana, mudah digunakan, memiliki kekuatan dalam melatih data yang mempunyai banyak noise, dan efektif untuk proses pengelompokan[11].

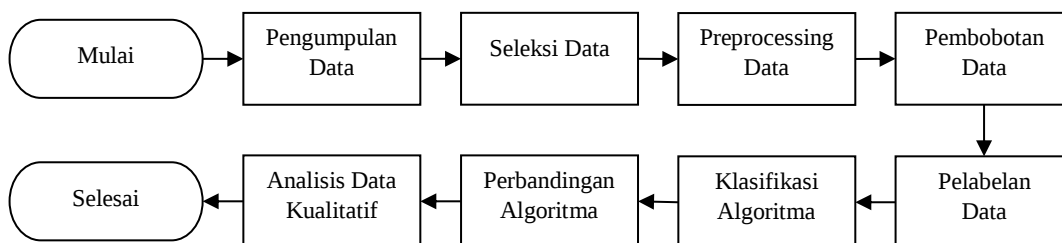
Perbandingan algoritma pada penelitian ini dilakukan untuk mengembangkan dan memilih model terbaik diantara kedua algoritma dalam klasifikasi teks. Sebagian besar peneliti telah berupaya melakukan penelitian untuk menemukan metode dan algoritma yang optimal dalam mengklasifikasikan analisis sentimen[14], selain itu kinerja yang beragam juga mendorong dilakukannya perbandingan algoritma pada penelitian ini. Hasil yang ingin dicapai dari perbandingan tersebut adalah untuk mendapatkan wawasan yang lebih baik tentang kinerja dari kedua model yang digunakan, seperti performa, apakah salah satu algoritma unggul secara signifikan dibandingkan yang lain atau memiliki kinerja yang sama.

Penelitian ini bertujuan menganalisis pandangan masyarakat terkait fitur TikTok Shop dengan memanfaatkan Twitter sebagai sumber data. Opini yang terkandung dalam tweet dianalisis melalui teknik text mining menggunakan algoritma NBC dan KNN. Proses pelabelan dilakukan dengan teknik clustering menggunakan metode K-Means, yang menghasilkan klasifikasi opini ke dalam tiga kategori, yaitu negatif, positif, dan netral. Kemudian dilakukan penambahan teknik analisis data kualitatif dengan metode thematic analysis bertujuan untuk menemukan pola atau tema dalam data.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah yang dilakukan dalam melakukan penelitian. Secara garis besar tahapan penelitian meliputi, pengumpulan data, preprocessing data, pembobotan data, pelabelan, klasifikasi algoritma yang digunakan, perbandingan algoritma, dan terakhir dilakukan analisis data kualitatif. Tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian



## 2.2 Pengumpulan Data

Data dalam penelitian ini merupakan teks yang diambil dari tweet di platform media sosial Twitter. Pencarian dilakukan dengan dua kata kunci yaitu "tiktok shop" dan "fitur" dalam bahasa Indonesia, dan rentang waktu pengumpulan data adalah mulai dari 13 September hingga 13 Oktober 2023. Metode pengumpulan data yang digunakan adalah teknik web scraping dengan memanfaatkan alat bantu Google Colab dan bahasa pemrograman python.

## 2.3 Seleksi Data

Proses seleksi data melibatkan penghapusan tweet yang tidak relevan dengan topik penelitian[15]. Sebagai contoh, seperti tweet yang digunakan untuk mempromosikan penjualan produk atau tweet yang berisi informasi berita dari berbagai situs yang dibagikan di platform Twitter.

## 2.4 Preprocessing Data

Preprocessing merupakan metode efektif untuk membersihkan dan mengorganisir data yang awalnya bersifat tidak terstruktur menjadi data yang terstruktur dan memiliki makna[16]. Preprocessing adalah langkah persiapan data yang diperlukan klasifikasi sentimen. Tahapan preprocessing dalam penelitian ini adalah:

1. Cleaning  
Data-data yang tidak lengkap, mengandung eror, dan tidak konsisten akan dibuang dari koleksi data
2. Tokenize  
Pada proses tokenize ini semua tanda baca, simbol, dan karakter bukan huruf akan dihapus, misal (;), ("), (\*), (/)
3. Normalisasi  
Teks yang mengandung singkatan atau simbol akan diubah menjadi teks standar
4. Stopword removal  
Proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata
5. Stemming  
Menghilangkan imbuhan pada suatu kata

## 2.5 Pembobotan Data

TF-IDF memberikan bobot kepada kata-kata dalam dokumen dengan menggabungkan dua konsep yaitu frekuensi kemunculan kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut. Bobot ini mencerminkan tingkat pentingnya suatu kata dalam suatu dokumen[17].

## 2.6 Pelabelan

Data yang telah dikumpulkan dari media sosial Twitter awalnya belum terlabel, sehingga diperlukan proses pelabelan agar data dapat dianalisis. Dalam penelitian ini pelabelan dilakukan menggunakan teknik clustering, penelitian serupa menerapkan teknik clustering dilakukan oleh Mustakim, dkk[18]. Pelabelan dalam penelitian ini dibagi menjadi tiga kategori, yaitu negatif, netral, dan positif.

## 2.7 Klasifikasi Algoritma Naïve Bayes Classifier

Naïve Bayes classifier (NBC) adalah metode klasifikasi berdasarkan probabilitas[19]. Ada dua proses penting Naive Bayes dalam melakukan klasifikasi, dan itu adalah pelatihan dan pengujian. Berikut adalah persamaan NBC:

$$P(C|X) = \frac{P(C|X)P(C)}{P(X)} \quad (1)$$

## 2.8 Klasifikasi Algoritma KNN

Algoritma KNN adalah metode untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak paling dekat dengan objek tersebut[20]. Dalam algoritma ini, tetangga terdekat dihitung berdasarkan nilai k yang menentukan jumlah tetangga terdekat yang akan dipertimbangkan. Dengan demikian, nilai k menentukan kelas dari titik data sampel[21]. Berikut persamaan KNN:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (2)$$

## 2.9 Perbandingan Algoritma

Perbandingan pada penelitian ini dilakukan dengan metode confusion matrix, parameter yang dievaluasi adalah, akurasi, presisi, dan recall. Akurasi adalah nilai kemiripan antara nilai prediksi dan nilai sebenarnya. Presisi adalah tingkat kebenaran antara informasi yang diminta pengguna dan yang diberikan oleh sistem. Recall adalah tingkat keberhasilan dari sistem dalam menemukan kembali suatu informasi[22].

## 2.10 Analisis Data Kualitatif

### 3. HASIL DAN PEMBAHASAN

**Tabel 1.** Dataset awal

| No    | Username    | Tweet                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | naim__x     | Aku suka apa Indonesia buat. Aku tak minat apa TikTok Shop buat dengan bagi voucher, sampai harga drop dari harga market. Masalah dia, kau akan bergantung dengan TikTok lepas tu. Kau 'technically' berkerja, bergantung hidup dengan TikTok. Yang tak bersama TikTok, dimusnahkan.                                                                                                                                                  |
| 2     | k03825868   | Gmorn ready jual akun tiktok murah nih bsa live,tiktok shop affiliate on (min 1.000 fols), tempel web di bio bsa email / no hp bsa ganti 8Ÿ™ª8Ÿ™%o NOMOR WA : <a href="https://t.co/Wa0nj9jaMJ">https://t.co/Wa0nj9jaMJ</a> TESTIMONI 648+ : <a href="https://t.co/TrqIe40vSj">https://t.co/TrqIe40vSj</a> wtb wts need jual acc akun tiktok murah #ZonaUang #ZonaJajan <a href="https://t.co/9DZS8RJkO8">https://t.co/9DZS8RJkO8</a> |
| 5.190 | sweethdevil | Nah ngapa juga minta dihapus... pedagang offline yg ngga bisa adaptasi ya bakalan begitu terus jualannya, padahal banyak kok pengusaha lokal yang sukses lewat tiktok shop                                                                                                                                                                                                                                                            |

**Tabel 2.** Hasil seleksi data

| No    | Username  | tweet                                                                                                                                                                 |
|-------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | namakuaji | Ada wacana TikTok Shop dilarang, Padahal banyak banget UMKM yang terbantu dengan hadirnya TikTok Shop                                                                 |
| 2     | kinandrst | @jspxroid Jujur gak setuju ka, menurut saya semua diuntungkan disini.. umkm jg bnyk yg terbantu kok.. kebetulan mau angkat kasus tiktok shop buat tesis saya ðŹŹ¹     |
| 2.339 | spcmonn   | @kegblgnunfaedh tiktok shop sebeaneernya ada plus minus sih, cuma emg algoritmanya aja kadang yg nyebelín. tapi kalau izinnyasudah diperbarui semoga lebih baik lagi. |

**Tabel 3.** Hasil cleaning

| No    | Cleaning                                                                                                                                          |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | ada wacana tiktok shop dilarang padahal banyak banget umk yang terbantu dengan hadirnya tiktok shop                                               |
| 2     | jujur gak setuju ka menurut saya semua diuntungkan disini umkm jg bnyk yg terbantu kok kebetulan mau angkat kasus tiktok shop buat tesis saya     |
| ....  | ....                                                                                                                                              |
| 2.339 | tiktok shop sebeneernya ada plus minus sih cuma emg algoritmanya aja kadang yg nyebelin tapi kalau izinnyasudah diperbarui semoga lebih baik lagi |

## Submitted: 24/12/2023; Accepted: 09/01/2024; Published: 10/01/2024



Tabel 4. Hasil Tokenizing

| No    | Tokenizing                                                                                                                                                                                                             |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | ['ada', 'wacana', 'tiktok', 'shop', 'dilarang', 'padahal', 'banyak', 'banget', 'umk', 'yang', 'terbantu', 'dengan', 'hadirnya', 'tiktok', 'shop']                                                                      |
| 2     | ['jujur', 'gak', 'setuju', 'ka', 'menurut', 'saya', 'semua', 'diuntungkan', 'disini', 'umkm', 'jg', 'bnyk', 'yg', 'terbantu', 'kok', 'kebetulan', 'mau', 'angkat', 'kasus', 'tiktok', 'shop', 'buat', 'tesis', 'saya'] |
| 2.339 | ['tiktok', 'shop', 'sebeneernya', 'ada', 'plus', 'minus', 'sih', 'cuma', 'emg', 'algoritmanya', 'aja', 'kadang', 'yg', 'nyebelin', 'tapi', 'kalau', 'izinnyasudah', 'diperbarui', 'semoga', 'lebih', 'baik', 'lagi']   |

### 3.3.3 Normalisasi

Normalisasi teks yang digunakan pada penelitian ini untuk mengubah teks asli yang berisi singkatan dan simbol menjadi teks yang standar. Hasil normalisasi dapat dilihat pada tabel 5.

Tabel 5. Hasil normalisasi

| No    | Normalisasi                                                                                                                                                                                                                      |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | ['ada', 'wacana', 'tiktok', 'shop', 'dilarang', 'padahal', 'banyak', 'banget', 'umk', 'yang', 'terbantu', 'dengan', 'hadirnya', 'tiktok', 'shop']                                                                                |
| 2     | ['jujur', 'enggak', 'setuju', 'kak', 'menurut', 'saya', 'semua', 'diuntungkan', 'disini', 'umkm', 'juga', 'banyak', 'yang', 'terbantu', 'kok', 'kebetulan', 'mau', 'angkat', 'kasus', 'tiktok', 'shop', 'buat', 'tesis', 'saya'] |
| 2.339 | ['tiktok', 'shop', 'sebeneernya', 'ada', 'plus', 'minus', 'sih', 'hanya', 'memang', 'algoritmanya', 'saja', 'kadang', 'yang', 'menyebalkan', 'tapi', 'kalau', 'izinnyasudah', 'diperbarui', 'semoga', 'lebih', 'baik', 'lagi']   |

### 3.3.4 Remove Stopword

Remove Stopword adalah penghapusan kata berdasarkan kata yang terdapat dalam stoplist. Hasil dari Remove Stopword dapat dilihat pada tabel 6.

Tabel 6. Hasil Remove stopwords

| No    | Remove Stopword                                                                                                                            |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | ['wacana', 'tiktok', 'shop', 'dilarang', 'banget', 'umk', 'terbantu', 'hadirnya', 'tiktok', 'shop']                                        |
| 2     | ['jujur', 'setuju', 'kak', 'diuntungkan', 'umkm', 'terbantu', 'angkat', 'tiktok', 'shop', 'tesis']                                         |
| 2.339 | ['tiktok', 'shop', 'sebeneernya', 'plus', 'minus', 'sih', 'algoritmanya', 'kadang', 'menyebalkan', 'izinnyasudah', 'diperbarui', 'semoga'] |

### 3.3.5 Stemming

Proses penguraian kata menjadi bentuk kata dasar. Hasil dari Stemming dapat dilihat pada tabel 7.

Tabel 7. Hasil Stemming

| No    | Stemming                                                                                                                     |
|-------|------------------------------------------------------------------------------------------------------------------------------|
| 1     | ['wacana', 'tiktok', 'shop', 'larang', 'banget', 'umk', 'bantu', 'hadir', 'tiktok', 'shop']                                  |
| 2     | ['jujur', 'tuju', 'kak', 'untung', 'umkm', 'bantu', 'angkat', 'tiktok', 'shop', 'tesis']                                     |
| ....  | ....                                                                                                                         |
| 2.339 | ['tiktok', 'shop', 'sebeneernya', 'plus', 'minus', 'sih', 'algoritmanya', 'kadang', 'sebal', 'izinnyasudah', 'baru', 'moga'] |

## 3.4 TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) dilakukan untuk mendapatkan bobot atau nilai dari suatu kata. Hasil TF-IDF dapat dilihat pada tabel 8.

Tabel 8. Hasil TF-IDF

| No    | tiktok   | shop     | umkm     | larang  | algoritma | bantu    |
|-------|----------|----------|----------|---------|-----------|----------|
| 1     | 0,215877 | 0,226786 | 0        | 0,32838 | 0         | 0,313876 |
| 2     | 0,07904  | 0,083034 | 0,212754 | 0       | 0         | 0,229841 |
| ....  | ....     | ....     | ....     | ....    | ....      | ....     |
| 2.339 | 0,058284 | 0,061229 | 0        | 0       | 0         | 0        |

## 3.5 Pelabelan Data

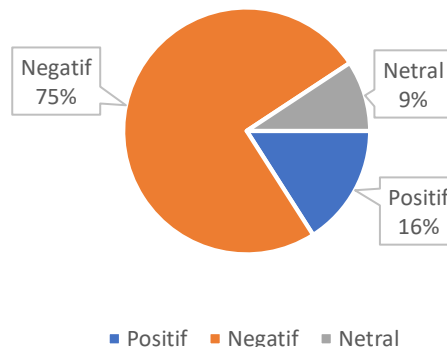
Data yang telah diperoleh dan telah dilakukan preprocessing tidak memiliki label, sehingga pada tahap pelabelan ini dilakukan dengan menggunakan algoritma K-Means Clustering dengan nilai K= 3. Hasil pelabelan data dapat dilihat pada tabel 9.



**Tabel 9.** Hasil pelabelan

| No    | Tweet                                                                      | Sentimen |
|-------|----------------------------------------------------------------------------|----------|
| 1     | wacana tiktok shop larang banget umk bantu hadir tiktok shop               | Netral   |
| 2     | jujur tuju kak untung umkm bantu angkat tiktok shop tesis                  | Negatif  |
| ....  | ....                                                                       | ....     |
| 2.339 | ngapa hapus dagang offline adaptasi ya jual usaha lokal sukses tiktok shop | Negatif  |

Dari hasil pelabelan data review terdapat sentimen positif sebanyak 373, pada sentimen negatif yaitu 1.749 dan sentimen netral sebanyak 217. Persentase data yang sudah diberi label dapat dilihat pada Gambar 2.



**Gambar 2.** Hasil pelabelan

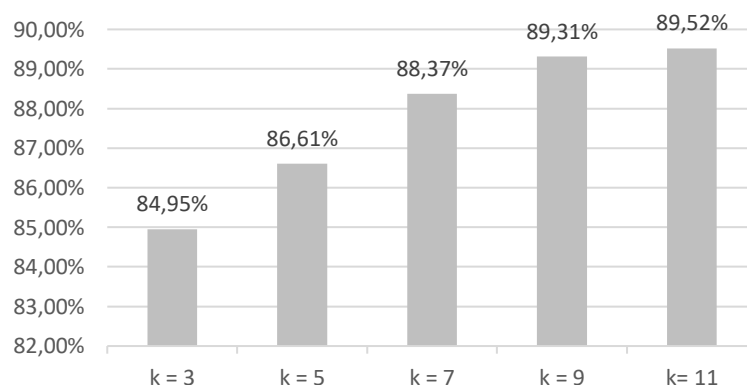
Dari analisis Gambar 2, dapat disimpulkan bahwa persentase tertinggi data terdapat pada sentimen negatif, mencapai 75%. Untuk sentimen menengah, persentasinya terdapat pada sentimen positif, yakni sebesar 16%, sementara sentimen netral memiliki persentase terendah, yaitu 9%.

### 3.6 Algoritma Naïve Bayes Classifier

Hasil percobaan dengan menerapkan metode 10 K-fold cross-validation pada dataset yang terdiri dari 2.339 data menunjukkan angka akurasi sebesar 71,05%, presisi mencapai 52,24%, dan recall sebesar 52,28%. Hasil ini memberikan gambaran tentang performa model dalam mengklasifikasikan data, di mana akurasi mencerminkan tingkat kebenaran klasifikasi secara keseluruhan, presisi menggambarkan seberapa tepat model dalam mengidentifikasi positif, dan recall mencerminkan kemampuan model dalam menemukan semua kasus positif yang sebenarnya.

### 3.7 Algoritma KNN

Dalam penelitian ini, algoritma KNN diuji sebanyak lima kali percobaan dengan variasi nilai parameter K, yaitu 3, 5, 7, 9, dan 11. Parameter yang dipilih untuk evaluasi adalah parameter yang menghasilkan nilai akurasi tertinggi. Berikut hasil percobaan nilai parameter K yang telah ditentukan pada Gambar 3.



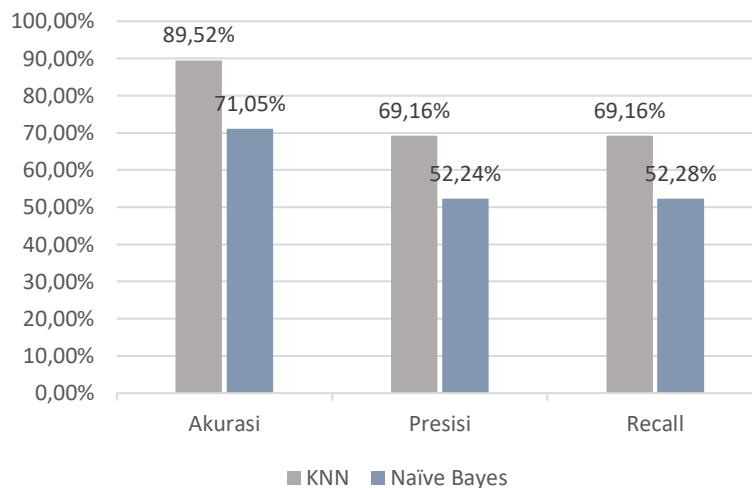
**Gambar 3.** Tingkat akurasi percobaan KNN

Hasil pengujian menunjukkan bahwa tingkat akurasi tertinggi tercapai pada parameter k=11, dengan nilai akurasi mencapai 89,52%. Sedangkan tingkat akurasi terendah terletak pada parameter k = 3 dengan tingkat akurasi 84,95%.



### 3.8 Perbandingan Algoritma

Melalui uji coba yang telah dilakukan, diperoleh hasil akurasi pada algoritma NBC sebesar 71,05%, presisi 52,24%, dan recall 52,28%. Sedangkan pada algoritma KNN menunjukkan hasil akurasi tertinggi sebesar 89,52%, presisi 69,16%, dan recall 69,16% terhadap data yang diperoleh. Dengan demikian, dapat dilihat bahwa akurasi yang dihasilkan menggunakan metode KNN lebih baik dibandingkan dengan NBC. Perbandingan performa dari kedua algoritma dapat dilihat pada Gambar 4.



Gambar 4. Perbandingan Algoritma NBC dan KNN

### 3.9 Analisis Data Kualitatif

Tujuan analisis data kualitatif pada penelitian ini adalah untuk menganalisa data dengan tujuan mengidentifikasi pola atau untuk menemukan tema melalui data yang telah dikumpulkan oleh peneliti (Braun & Clarke, 2006). Metode yang digunakan dalam penelitian ini adalah thematic analysis. Dalam hal ini setelah data dikumpulkan dilakukan prosedur pengkodean (coding) untuk menemukan tema-tema berdasarkan kata-kata atau frasa yang sering muncul dalam data, kemudian menghitung frekuensi munculnya tema-tema yang telah ditemukan dalam data. Hasil thematic analysis dapat dilihat pada tabel 10.

Tabel 10. Hasil thematic analysis

| No | Tema                       | Pernyataan                                                                                                                                                                                                                             | Koding                                                           | Frekuensi |
|----|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|-----------|
| 1  | Kemudahan berbelanja       | Udah lah gak usah ditutup, tiktokshop tuh membantu bgt tau karena murah + gak ada biaya admin / biaya layanan blablabla kaya si oren dan si ijo                                                                                        | Murah, Efisien, Cepat, nyaman, praktis                           | 571       |
| 2  | Algoritma fitur yang buruk | Tapi untuk tiktok shop yg bikin ngeri algoritmanya, ketika barang itu laku keras data yg dimiliki tiktok bisa diambil pihak china dengan langsung produksi besar yg otomatis harga jualnya lebih murah, dan itu yg mematikan umkm kita | Algoritma, predatory pricing, mematikan, monopoli, merusak pasar | 456       |
| 3  | Fitur tidak memuaskan      | Tiktok shop apa bagusnyah sih? Gw berkali-kali beli disitu, kualitasnya barangnya parah banget                                                                                                                                         | Jelek, parah, mahal, lambat, kualitas jelek                      | 130       |
| 4  | Regulasi tidak sesuai      | Daripada menutup tiktok shop, mending bikin umkm kita melek teknologi. Bikin regulasi yang meringankan umkm supaya harga bs bersaing dgn produk import                                                                                 | Regulasi, izin, aturan, kebijakan, aplikasi terpisah             | 754       |
| 5  | Fitur Bermanfaat           | padahal dengan adanya tiktok shop ini banyak yang merasakan terbantu untuk jualan                                                                                                                                                      | Membantu, menguntungkan, penghasilan, meningkatkan, pekerjaan    | 428       |

Melalui analisis data, ditemukan terdapat lima tema utama. Tema yang paling dominan adalah "Regulasi yang tidak sesuai" dengan total 754, mencakup sekitar 32% dari seluruh data. Hal ini menunjukkan bahwa tiktok



shop tidak memiliki izin atas usaha dan menyalahi aturan yang telah ditetapkan. Tema selanjutnya yang cukup dominan ketiga adalah “algoritma yang buruk”, algoritma pada fitur TikTok Shop kurang adil, karena manipulasi algoritma ini memberikan keuntungan lebih kepada satu produk, sementara pada saat yang sama merugikan produk lainnya. Ini sejalan dengan pernyataan Staf Khusus Menteri koperasi dan UKM (MenKopUKM) Bidang Pemberdayaan Ekonomi Kreatif, Fiki Satari. Hasil thematic analysis ini dapat memberikan manfaat untuk perbaikan fitur atau aplikasi dimasa mendatang seperti, 1. Memahami kebutuhan, harapan, dan masalah pelanggan terkait fitur tiktok shop, sehingga dapat meningkatkan kualitas layanan dan kepuasan pelanggan untuk pengembangan aplikasi dimasa mendatang, 2. Mengetahui manfaat dan dampak dari fitur tiktok shop, menurut analisis data yang dilakukan tiktok shop sendiri sangat membantu sebagian para pedagang kecil untuk membangun usahanya menjadi lebih baik bukan hanya itu fitur tiktok shop sendiri banyak menciptakan lapangan pekerjaan seperti, jumlah kurir paket yang bertambah dengan adanya tiktok shop, host, konten kreator, editor, dan lain-lain. 3. Mendapatkan ide dan inspirasi untuk mengembangkan suatu aplikasi tiktok shop yang lebih baik, inovatif, dan sesuai dengan tren dan permintaan pasar. Dengan analisis data ini, dapat mengenali pola-pola penggunaan yang bermanfaat untuk merancang strategi pembuatan aplikasi yang lebih efisien dan juga mengadaptasi produk sesuai dengan keinginan pengguna.

### 3.10 Pembahasan

Berdasarkan hasil yang diperoleh percobaan diatas diketahui bahwa algoritma KNN lebih baik dibandingkan naïve bayes, hal ini sejalan dengan penelitian yang dilakukan oleh Hilman Wisnu, dkk, dimana KNN memiliki akurasi yang lebih baik daripada Naive Bayes[11]. Akan tetapi penelitian yang dilakukan oleh Sevira Wongkar, 2019 adalah sebaliknya, dimana Penelitian tersebut menemukan bahwa algoritma Naïve Bayes Classifier memiliki tingkat akurasi lebih tinggi daripada algoritma KNN[6].

## 4. KESIMPULAN

Penelitian ini bertujuan untuk melakukan analisis sentimen menggunakan dua algoritma, yaitu Naïve Bayes Classifier (NBC) dan K-Nearest Neighbour (KNN), dengan menerapkan metode validasi K-Fold Cross Validation yang memiliki nilai  $k=10$ . Data awal yang diperoleh berjumlah 5.190, namun dilakukan seleksi data untuk mendapatkan data yang relevan dengan topik penelitian, sehingga diperoleh 2.339 data. Proses pelabelan dilakukan menggunakan teknik clustering dengan algoritma K-Means dan nilai  $k=3$ . Pada algoritma KNN, dilakukan uji coba dengan lima parameter K berbeda, yaitu 3, 5, 7, 9, dan 11. Selanjutnya, perbandingan performa kedua algoritma dilihat dari nilai akurasi tertinggi yang dihasilkan. Hasil akurasi pada algoritma NBC adalah sebesar 71,05%, sementara algoritma KNN menunjukkan hasil akurasi tertinggi sebesar 89,52%. Dari perbandingan ini, dapat disimpulkan bahwa model KNN lebih unggul dibandingkan dengan NBC dalam melakukan analisis sentimen terhadap data yang telah diolah. Hasil thematic analysis mendapatkan total ada lima tema dalam data dan yang paling dominan adalah tema “Regulasi yang tidak sesuai”. Hasil ini dapat dijadikan oleh platform tiktok untuk memperbaiki dengan menertibkan regulasi atas fitur tersebut supaya fitur tersebut dapat diterima dengan baik oleh masyarakat dan tidak dianggap menyalahi aturan, serta memperbaiki algoritma menjadi lebih baik, tidak menurunkan traffic penjualan dari produk lokal, dan menetapkan harga dari suatu produk tidak melampaui terlalu jauh dari pasaran supaya UMKM dapat bersaing secara sehat.

## REFERENCES

- [1] D. E. Oktania and T. A. Indarwati, “Pengaruh Perceived Usefulness, Perceived Ease Of Use, Dan Compatibility With Lifestyle Terhadap Niat Beli Di Social Commerce,” *Jurnal Ilmu Manajemen*, vol. 10, no. 1, pp. 255–267, Mar. 2022, doi: 10.26740/jim.v10n1.p255-267.
- [2] D. Amanah, D. Ansari Harahap, and M. Gunarto, “Perceived Risk and Security in Creating Online Purchasing Decision at Marketplace in Indonesia,” *Journal of Applied Business and Economics (JABE)*, vol. 7, no. 2, pp. 162–179, Dec. 2020, doi: 10.30998/jabe.v7i2.7553.
- [3] A. A. A. Sharabati, S. Al-Haddad, M. Al-Khasawneh, N. Nababteh, M. Mohammad, and Q. Abu Ghoush, “The Impact of TikTok User Satisfaction on Continuous Intention to Use the Application,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 3, Sep. 2022, doi: 10.3390/joitmc8030125.
- [4] I. Purwandani, N. O. Syamsiah, and S. Nurwahyuni, “Perceived Usability Evaluation of TikTok Shop Platform Using the System Usability Scale,” *Sinkron*, vol. 8, no. 3, pp. 1389–1399, Jul. 2023, doi: 10.33395/sinkron.v8i3.12473.
- [5] S. Septyaningsih, M. T. Multazam, and B. Sobirov, “Legal Protection of Consumer Rights in Transactions at TikTok Shop: Unraveling New Legal Insights,” *Kosmik Hukum*, vol. 23, no. 3, p. 248, Aug. 2023, doi: 10.30595/kosmikhukum.v23i3.17396.
- [6] M. Wongkar and A. Angdresy, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter,” 2019 Fourth International Conference on Informatics and Computing (ICIC), Oct. 2019, doi: 10.1109/ICIC47613.2019.8985884.
- [7] F. Aftab et al., “A Comprehensive Survey on Sentiment Analysis Techniques,” *International Journal of Technology*, vol. 14, no. 6, pp. 1288–1298, 2023, doi: 10.14716/ijtech.v14i6.6632.
- [8] C. Sindhu, B. Sasmal, R. Gupta, and J. Prathipa, “Subjectivity detection for sentiment analysis on Twitter data,” *Lecture Notes in Networks and Systems*, vol. 130, pp. 467–476, 2021, doi: 10.1007/978-981-15-5329-5\_43.



- [9] A. J. Nair, G. Veena, and A. Vinayak, "Comparative study of Twitter Sentiment on COVID - 19 Tweets," in Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 1773–1778. doi: 10.1109/ICCMC51019.2021.9418320.
- [10] A. R. Atmadja, W. Uriawan, F. Pritisen, D. S. Maylawati, and A. Arbain, "Comparison of Naive Bayes and K-nearest neighbours for online transportation using sentiment analysis in social media," in Journal of Physics: Conference Series, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1402/7/077029.
- [11] H. Wisnu, M. Afif, and Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," in Journal of Physics: Conference Series, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1742-6596/1444/1/012034.
- [12] H. Sudira, A. L. Diar, and Y. Ruldeviyani, "Instagram Sentiment Analysis with Naive Bayes and KNN Exploring Customer Satisfaction of Digital Payment Services in Indonesia," 2019 International Workshop on Big Data and Information Security (IWBIS), Oct. 2019, doi: 10.1109/IWBIS.2019.8935700.
- [13] E. Utami, S. Raharjo, A. Dwi Hartanto, S. Adi, and A. Noor Ichsan, "K-Nearest Neighbor and Naive Bayes Classifier Comparison for Individual Character Classification on Twitter," in 2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ICORIS50180.2020.9320759.
- [14] F. A. Wenando, R. Hayami, Bakaruddin, and A. Y. Novermahakim, "Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms," in Proceeding - 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, ICITAMEE 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 279–282. doi: 10.1109/ICITAMEE50454.2020.9398513.
- [15] S. D. Pramukti, A. Nugroho, and A. S. Sunge, "Analisis Sentimen Masyarakat Dengan Metode Naive Bayes dan Particle Swarm," Techno.COM, vol. 21, no. 1, pp. 62–75, Feb. 2022, doi: 10.33633/tc.v21i1.5332.
- [16] S. Pradha, M. N. Halgamuge, and T. Q. V. Nguyen, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data," 2019 11th International Conference on Knowledge and Systems Engineering (KSE), 2019, pp. 1–8. doi: 10.1109/KSE.2019.8919368.
- [17] Sintia, S. Defit, and G. W. Nurcahyo, "Product Codification Accuracy With Cosine Similarity And Weighted Term Frequency And Inverse Document Frequency (Tf-Idf)," Journal of Applied Engineering and Technological Science, vol. 2, no. 2, pp. 14–21, May 2021, doi: 10.37385/jaets.v2i2.210.
- [18] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," in Journal of Physics: Conference Series, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1783/1/012016.
- [19] Suprianto, M. Fadlan, Muhammad, Y. Amaliah, and Mussallimah, "Retrieval information using generalized vector space models and sentiment analysis using naïve bayes classifier for evaluation of lecturers by students," in 2020 5th International Conference on Informatics and Computing, ICIC 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICIC50835.2020.9288584.
- [20] F. Firmansyah et al., "Comparing Sentiment Analysis of Indonesian Presidential Election 2019 with Support Vector Machine and K-Nearest Neighbor Algorithm," in 6th International Conference on Computing, Engineering, and Design, ICCED 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ICCED51276.2020.9415767.
- [21] I. A. Abu Amra and A. Y. A. Maghari, "Students Performance Prediction Using KNN and Naïve Bayesian," 8th International Conference on Information Technology (ICIT), 2017, pp. 909–913. doi: 10.1109/ICITECH.2017.8079967.
- [22] A. R. Isnain, H. Sulistiani, B. M. Hurohman, A. Nurkholis, and Styawati, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 8, no. 2, Aug. 2022, doi: 10.26418/jp.v8i2.54704.