

# Klasifikasi Ulasan Aplikasi TikTok Menggunakan Algoritma K-Nearest Neighbor dan Chi Square

Sandrina Ferani Aisyah Putri<sup>a1</sup>, I Wayan Supriana<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Udayana, Bali  
Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 08261, Bali, Indonesia  
<sup>1</sup>sandrinaferani060@student.unud.ac.id  
<sup>2</sup>wayan.supriana@unud.ac.id

## Abstract

*TikTok application has achieved extraordinary popularity among users around the world, which has been downloaded by more than 500 million users with 16 million reviews and received a rating of 4.4 out of 5 on the Google Play Store. In this study, we will analyze user sentiment towards the TikTok application reviews. These reviews can be a benchmark for users to find out information about user experience and become a race for application developers to improve performance or quality. For that we need a method to describe the reviewer efficiently so that it is easier to understand the reviewer. In this study, the authors used a comparison of the KNN algorithm with the effect of feature selection to carry out the classification. Classification of application reviews into two classes, positive reviews, and negative reviews. In this classification, it is found that using Chi Square feature selection can produce the highest accuracy, with  $k = 9$  value of 86.22% whereas without Chi Square feature selection it only produces the highest accuracy with  $k = 11$  value of 77.04%.*

**Keywords:** TikTok, Classifier, Analysis Sentiment, K-Nearest Neighbor, Chi Square

## 1. Pendahuluan

Dalam beberapa tahun terakhir, aplikasi TikTok telah meraih popularitas yang luar biasa di kalangan pengguna di seluruh dunia. TikTok memungkinkan pengguna untuk membuat, menonton, membagikan, dan berinteraksi dengan beragam konten kreatif, seperti tarian, seputar informasi, tips dan trik, komedi, tantangan, lip-sync, ataupun hal lain sesuai minat pengguna. Bahkan sejak September 2021, TikTok sudah menyediakan fitur *social commerce* yang memungkinkan pengguna dan kreator untuk mempromosikan serta menjual produknya melalui TikTok Shop.

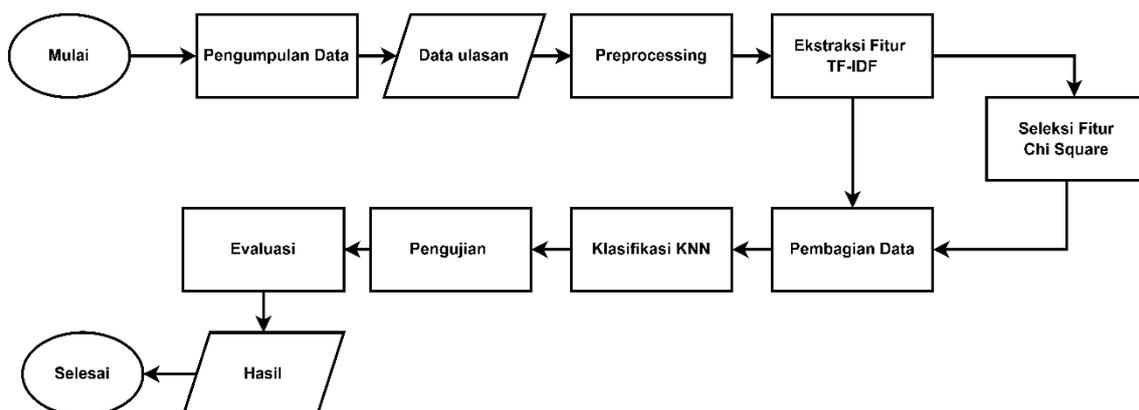
Indonesia menempati peringkat kedua dengan jumlah pengguna TikTok terbanyak dunia, yakni mencapai 109,9 juta pengguna pada bulan Januari 2023 menurut sumber databoks.katadata.co.id [1]. Peningkatan popularitas TikTok adalah indikasi positif bahwa masyarakat sudah lebih beradaptasi dengan perkembangan teknologi. Selain itu, aplikasi TikTok sudah diunduh oleh lebih dari 500 juta lebih pengguna dengan 17 juta ulasan dan mendapatkan rating 4,4 dari 5 pada Google Play Store. Meski memiliki rating yang tinggi, tetapi tidak semua ulasan tersebut memiliki arti yang positif, melainkan juga negatif. Ulasan-ulasan tersebut dapat menjadi tolak ukur bagi pengguna untuk mengetahui informasi mengenai pengalaman pengguna baik dari kelebihan atau kekurangan aplikasi dan juga menjadi tolak ukur bagi perusahaan untuk mengembangkan serta meningkatkan kualitas aplikasi. Oleh karena itu, dalam mengelompokkan banyaknya ulasan pengguna dengan mudah dan cepat, dapat menggunakan metode klasifikasi.

Penelitian sebelumnya mengenai ulasan aplikasi TikTok pernah dilakukan pada tahun 2021 dengan judul "Analisis Sentimen Ulasan Aplikasi TikTok di Google Play Menggunakan Metode Support Vector Machine (SVM) dan Asosiasi" [2]. Penelitian tersebut mengambil ulasan data dari bulan September 2020 hingga Februari 2021 sebanyak 3200 ulasan dengan jumlah ulasan positif sebanyak 1741 dan jumlah ulasan negative sebanyak 1459 yang menghasilkan tingkat *accuracy*

dan kappa terbaik sebesar 90,62% dan 81,24%. Penelitian lainnya dilakukan pada tahun 2021 juga dengan judul “Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Propotional Difference (CPD)” dengan menggunakan 1000 data dan memiliki dua kelas yang menghasilkan nilai *accuracy* sebesar 0,729947, *precision* sebesar 0,746854, *recall* sebesar 0,926118, dan *nilai f-measure* 0,824511 [3]. Tentunya data ulasan tersebut telah tidak relevan lagi saat ini sehingga penulis mengambil topik ini. Perbedaan penelitian ini dengan penelitian sebelumnya adalah terletak pada metode algoritma yang digunakan, yaitu menggunakan metode K-Nearest Neighbor dan seleksi fitur Chi Square dengan dataset yang berbeda juga. Hasil akhir dari penelitian ini adalah tingkat presentase performa berdasarkan jumlah ulasan aplikasi TikTok yang telah diklasifikasikan menjadi kelompok positif dan negatif.

## 2. Metode Penelitian

Pada tahap ini digambarkan alur penelitian yang akan dilakukan oleh penulis. Input pada penelitian ini adalah berupa ulasan aplikasi Tik Tok, kemudian data dibagi menjadi data latih dan data uji serta dilabeli oleh dua kelas, yaitu berupa ulasan positif dan ulasan negatif. *Output* yang akan dihasilkan adalah hasil evaluasi dari algoritma KNN tanpa seleksi fitur dengan algoritma KNN menggunakan seleksi fitur. Alur penelitian ini bisa dilihat pada gambar 1, ada dua alur penelitian yang dilakukan, alur pertama yang dilakukan, yaitu melakukan proses pengumpulan data, melakukan *preprocessing*, melakukan pembobotan fitur menggunakan ekstraksi fitur TF-IDF, melakukan penyeleksian fitur menggunakan Chi Square, melakukan klasifikasi menggunakan algoritma KNN, kemudian diuji menggunakan berbagai parameter k sehingga menghasilkan hasil evaluasi berdasarkan *precision*, *recall*, *F-1 score*, dan juga akurasi.



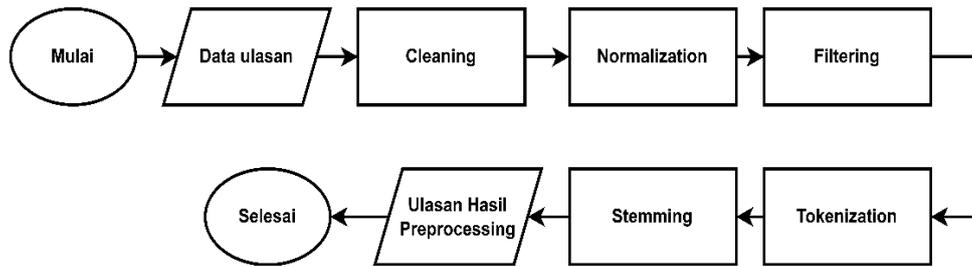
Gambar 1. Alur Penelitian

### 2.1 Pengumpulan Data

Pengumpulan data diambil berasal dari situs website Google Play Store dengan metode Scraping. Data yang dipilih adalah *review* atau ulasan pengguna berbahasa Indonesia dari aplikasi Tiktok yang dimulai dari periode bulan Januari hingga Juni 2023 dan disimpan dalam bentuk (.csv). Data ulasan yang dipakai berjumlah 1960 ulasan, yaitu 980 ulasan positif dan 980 ulasan negative.

### 2.2 Preprocessing

*Text preprocessing* adalah tahapan awal pemrosesan teks untuk menyiapkan data teks menjadi data yang siap diolah dalam tahap selanjutnya [4]. Gambar berikut adalah alur dari *preprocessing*.



**Gambar 2.** Alur *Preprocessing*

a. *Cleaning*

*Cleaning* adalah tahap membersihkan dokumen dari *noise*, seperti penghilangan duplikasi kata, penghilangan angka, karakter selain huruf alfabet, tanda baca seperti titik, koma, tanda tanya, tanda seru, dan lainnya. Pada tahap ini juga melakukan case folding, yaitu mengganti seluruh huruf yang ada pada dokumen teks menjadi *lowercase*. Hasil proses *cleaning* dapat dilihat pada tabel 1.

**Tabel 1.** Hasil *Cleaning*

<b>Data</b>	<b>Cleaning</b>
KENAPA YA SAYA GA BISA BIKIN STORY ATAU LIHAT STORY, PADAHAL UDH SAYA DOWNLOAD ULANG. TOLONG JAWABANNYA	kenapa ya saya ga bisa bikin story atau lihat story padahal udh saya download ulang tolong jawabannya

b. *Normalization*

*Normalization* adalah tahap menormalisasi bahasa pada bahasa yang tidak baku, misal bahasa gaul. Tujuannya adalah mengembalikan penulisan pada kata tersebut menjadi ke bentuk kata yang sesuai dengan KBBI. Hasil proses *normalization* dapat dilihat pada tabel 2.

**Tabel 2.** *Normalization*

<b>Data Cleaning</b>	<b>Normalization</b>
kenapa ya saya ga bisa bikin story atau lihat story padahal udh saya download ulang tolong jawabannya	kenapa ya saya tidak bisa bikin cerita atau lihat cerita padahal sudah saya unduh ulang tolong jawabannya

c. *Filtering*

*Filtering* adalah tahap menghapus kata-kata umum, seperti kata kata penghubung, waktu, dan lain sebagainya. Hasil proses *filtering* dapat dilihat pada tabel 3.

**Tabel 3.** *Filtering*

<b>Data Normalization</b>	<b>Filtering</b>
kenapa ya saya tidak bisa bikin cerita atau lihat cerita padahal sudah saya unduh ulang tolong jawabannya	ya tidak bikin cerita lihat cerita padahal unduh ulang tolong jawabannya

d. *Tokenization*

*Tokenization* adalah tahap pembagian atau pemisahan kata dalam suatu kalimat sehingga terbagi menjadi token-token tertentu. Hasil dari proses *tokenization* dapat dilihat pada tabel 4.

**Tabel 4. Tokenization**

<b>Data Filtering</b>	<b>Tokenization</b>
ya tidak bikin cerita lihat cerita padahal unduh ulang tolong jawabannya	['ya', 'tidak', 'bikin', 'cerita', 'lihat', 'cerita', 'padahal', 'unduh', 'tolong', 'jawabannya']

e. *Stemming*

*Stemming* adalah tahap memperoleh kata dasar dari suatu token dengan cara menghilangkan awalan, akhiran, sisipan, dan kombinasi dari awalan maupun akhiran. Hasil proses *stemming* dapat dilihat pada tabel 5.

**Tabel 5. Stemming**

<b>Data Tokenization</b>	<b>Stemming</b>
['ya', 'tidak', 'bikin', 'cerita', 'lihat', 'cerita', 'padahal', 'unduh', 'tolong', 'jawabannya']	['ya', 'tidak', 'bikin', 'cerita', 'lihat', 'cerita', 'padahal', 'unduh', 'tolong', 'jawaban']

**2.3 Ekstraksi Fitur Term Frequency Inverse Document Frequency (TF-IDF)**

TF-IDF merupakan suatu metode pembobotan kata yang dikenal baik dalam mengevaluasi pentingnya sebuah kata yang ada dalam dokumen [5]. Ekstraksi fitur atau pemberian bobot pada setiap kata bertujuan untuk mengetahui seberapa sering suatu kata muncul didalam dokumen. Tahapan untuk melakukan pembobotan kata dengan TF-IDF sebagai berikut:

- a. Hitung *term frequency* (tf).

$$tf(t, d) = \left( \frac{f(t,d)}{n} \right) \tag{1}$$

- b. Hitung IDF

$$idf(t) = \log \left( \frac{n}{1+df(t)} \right) \tag{2}$$

- c. Hitung bobot TF-IDF

$$w_{t,d} = tf_{t,d} \times idf_t \tag{3}$$

Keterangan :

- $w_{i,j}$  = bobot *term* i terhadap dokumen j
- $tf_{i,j}$  = frekuensi *term* i pada dokumen j
- $idf_i$  = nilai bobot IDF *term* i
- $N$  = jumlah dokumen keseluruhan
- $df_i$  = jumlah dokumen yang mengandung *term* i

**2.4 Seleksi Fitur Chi Square**

Seleksi fitur merupakan penghilangan sejumlah fitur-fitur yang tidak ada hubungannya dengan kategori pada dokumen sehingga tujuannya ialah memilih fitur-fitur yang penting dan relevan pada dokumen untuk proses kategorisasi [6].

$$X^2(t, c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \tag{4}$$

Keterangan:

- $N$  = jumlah dokumen keseluruhan
- $A$  = jumlah dokumen yang memiliki term  $t$  pada kelas  $c$ ,
- $B$  = jumlah dokumen yang memiliki term  $t$  tetapi tidak ditemukan pada kelas  $c$ ,

- $C$  = jumlah dokumen yang tidak memiliki term  $t$  pada kelas  $c$ ,
- $D$  = jumlah dokumen yang tidak memiliki term  $t$  tetapi tidak ditemukan pada kelas  $c$

## 2.5 Pembagian Data

Sebelum melakukan tahapan klasifikasi metode K-Nearest Neighbors perlu dilakukan pembagian data, baik setelah seleksi fitur atau setelah ekstraksi fitur. Dataset sentimen yang digunakan dilabeli dengan dua kelas yaitu berupa ulasan atau *review* yang positif dan ulasan atau *review* yang negatif, ulasan yang memiliki rating bintang 4 dan 5 merupakan ulasan positif sedangkan rating bintang 1 dan 2 merupakan ulasan negatif. Kemudian dilakukan tahap pembagian dataset menjadi data latih dan data uji. Data latih yang digunakan adalah sebesar 90% data latih dan 10% data uji. Setelah dataset dibagi dilanjutkan pengklasifikasian dengan metode K-Nearest Neighbors.

## 2.6 Klasifikasi KNN (K-Nearest Neighbor)

K-Nearest Neighbor (KNN) adalah suatu metode klasifikasi yang termasuk algoritma supervised learning, yaitu bergantung pada dataset yang telah diberikan pelabelan sebelumnya sebagai data latih untuk menghasilkan klasifikasi secara otomatis [7]. Prinsip kerja k-Nearest Neighbor (k-NN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya dalam data pelatihan. Berikut urutan proses kerja k-NN [8]:

- a. Menentukan parameter k (jumlah tetangga paling dekat).
- b. Menghitung kuadrat jarak euclidean (euclidean distance) masing-masing obyek terhadap data sampel yang diberikan. Rumus K-Nearest Neighbors (KNN) dapat dilihat pada persamaan 5.

$$d_i = \sum_{i=1}^p (x_{2i} - x_{1i})^2 \quad (5)$$

Keterangan:

- $d$  = jarak
  - $x_1$  = data training atau data latih
  - $x_2$  = data testing atau data uji
  - $i$  = variabel pada data
  - $p$  = dimensi suatu data
- c. Kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak euclid terkecil
  - d. Mengumpulkan kategori Y (Klasifikasi nearest neighbor)

## 2.7 Pengujian

Pengujian dilakukan untuk mengetahui bagaimana kinerja sistem itu sendiri dalam melakukan tugasnya, yaitu mengklasifikasikan sentimen ulasan atau *review* pengguna. Pada penelitian ini pengujian yang akan dilakukan adalah dengan mengubah parameter k pada metode K-Nearest Neighbor saja dan pada metode K-Nearest Neighbor dengan seleksi fitur Chi Square untuk menghasilkan nilai akurasi yang tertinggi.

## 2.8 Evaluasi

Evaluasi dihasilkan setelah melakukan pengujian dengan menggunakan *Confusion matrix* (CM). *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi dengan cara membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang dipakai untuk mengukur kemampuan suatu metode klasifikasi [9].

**Tabel 6. Confusion Matrix**

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

- *True Positive* = total data yang positif yang terklasifikasi benar.
- *True Negative* = total data yang negatif yang terklasifikasi benar.
- *False Positive* = total data yang positif tetapi terklasifikasi salah.
- *False Negative* = total data yang negatif tetapi terklasifikasi salah.

Berdasarkan data Confusion Matrix kemudian dapat diperoleh nilai Accuracy, Precision, Recall, dan F1-Score menggunakan persamaan berikut.

$$Accuracy = \frac{TP+TN}{(TP+FP+TN+FN)} \tag{6}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{7}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{8}$$

$$F1-Score = 2 * \frac{precision*recall}{precision+recall} \tag{9}$$

### 3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan tujuan untuk mengklasifikasikan ulasan pengguna ke dalam ulasan positif atau ulasan negatif dengan menggunakan algoritma K-Nearest Neighbour (KNN) saja dan algoritma K-Nearest Neighbour (KNN) dengan seleksi fitur Chi-Square yang diimplementasikan dengan bahasa pemrograman Python. Penentuan prediksi sentimen dilakukan dengan mencari label terdekat berdasarkan jarak antara data testing dengan data training yang sudah ada dan menentukan kategori mayoritas dari tetangga tersebut. Penelitian ini sekaligus untuk mengetahui pengaruh hasil evaluasi yang didapat jika menggunakan seleksi fitur atau tidak terhadap perubahan parameter K algoritma K-Nearest Neighbor. Perubahan nilai k pada percobaan adalah k = 3, k = 5, k = 7, k = 9, k = 11, dan k=13

#### 3.1. Hasil Metode K-Nearest Neighbor

Setelah dilakukan tahapan pre-processing, ekstrasi fitur TF-IDF, dan pembagian data, tahap selanjutnya adalah proses klasifikasi data testing berdasarkan data latih yang menggunakan metode K-Nearest Neighbor (KNN). Proses klasifikasi sekaligus pengujian ini akan membandingkan dari berbagai parameter K. Hasil akurasi masing-masing parameter k dapat dilihat pada tabel 7.

**Tabel 7. Akurasi metode KNN**

Nilai K	Accuracy
K = 3	68,87%
K = 5	71,93%
K = 7	73,46%
K = 9	74,48%
K = 11	77,04%
K = 13	76,53%

Berdasarkan pada tabel 7, pemilihan nilai K berpengaruh pada performance KNN, semakin besar nilai K yang ditentukan, akurasi juga semakin meningkat. Namun, ketika nilai K mencapai 13, akurasi sedikit menurun yang mungkin disebabkan karena batas-batas antar class semakin kabur. Oleh karena itu, nilai K terbaik yang memberikan akurasi tertinggi dalam pengujian ini adalah K = 11 dengan akurasi sebesar 77.04%. Klasifikasi sentimen ini menghasilkan jumlah prediksi benar sebanyak 151 dan jumlah prediksi salah sebanyak 45. Hasil dari *precision*, *recall* dan *F-1 score* didapatkan berdasarkan persamaan (7), (8) dan (9) yang ditunjukkan pada tabel 8.

**Tabel 8.** Hasil *precision*, *recall*, *f-1 score* dengan nilai k=11

	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
<b>Ulasan Positif</b>	80%	69%	74%
<b>Ulasan Negatif</b>	75%	84%	79%
<b>Rata-Rata</b>	77%	77%	77%

### 3.2. Hasil Metode K-Nearest Neighbor dan Chi Square

Setelah dilakukan tahapan pre-processing, ekstraksi fitur TF-IDF, dan pembagian data, tahap selanjutnya adalah proses klasifikasi data testing berdasarkan data latih yang menggunakan metode K-Nearest Neighbor (KNN) dan seleksi fitur Chi Square. Proses klasifikasi sekaligus pengujian ini akan membandingkan dari berbagai parameter K. Hasil akurasi masing-masing parameter k dapat dilihat pada tabel 9.

**Tabel 9.** Akurasi metode KNN dan Chi Square

Nilai K	<i>Accuracy</i>
<b>K = 3</b>	79,59%
<b>K = 5</b>	82,65%
<b>K = 7</b>	83,67%
<b>K = 9</b>	86,22%
<b>K = 11</b>	83,16%
<b>K = 13</b>	82,14%

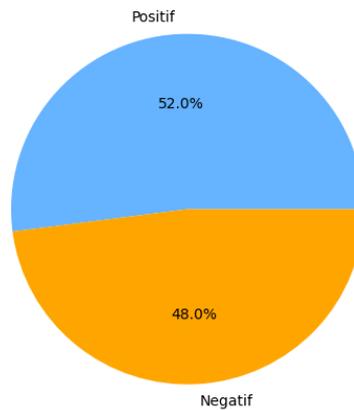
Berdasarkan pada tabel 9, pemilihan nilai K juga berpengaruh pada performance KNN dan Chi Square, semakin besar nilai K yang ditentukan, akurasi juga semakin meningkat. Namun, ketika nilai K mencapai 11, akurasi sedikit menurun. Oleh karena itu, nilai K terbaik yang memberikan akurasi tertinggi dalam pengujian ini adalah K = 9 dengan akurasi sekitar 86.22%. Klasifikasi sentimen ini menghasilkan jumlah prediksi benar sebanyak 169 dan jumlah prediksi salah sebanyak 27. Hasil dari *precision*, *recall* dan *F-1 score* didapatkan berdasarkan persamaan (7), (8) dan (9) yang ditunjukkan pada tabel 10.

**Tabel 10.** Hasil *precision*, *recall*, *f-1 score* dengan nilai k=9

	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
<b>Ulasan Positif</b>	85%	86%	86%
<b>Ulasan Negatif</b>	87%	86%	87%
<b>Rata-Rata</b>	86%	86%	86%

### 3.3. Hasil Klasifikasi Akurasi Tertinggi

Klasifikasi yang dilakukan pada 196 ulasan data testing dipilih yang menghasilkan akurasi tertinggi, yaitu pengujian menggunakan metode K-Nearest Neighbor, seleksi fitur Chi Square, K = 9.



**Gambar 3.** Hasil Klasifikasi

Berdasarkan visualisasi diagram pie, dapat dilihat hasil klasifikasi positif sebesar 52% yaitu sebanyak 102 ulasan, negatif sebesar 48% yaitu sebanyak 94 ulasan. Kata-kata yang sering muncul pada ulasan positif adalah bantu, ajar, asli. Sementara kata-kata yang sering muncul pada ulasan negatif adalah aneh, banyak, berita.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa implementasi seleksi fitur Chi Square dapat meningkatkan performa precision, recall, F1-Score, dan akurasi dibandingkan hanya dengan menggunakan metode K-Nearest Neighbor saja dalam mengklasifikasikan ulasan atau review pengguna aplikasi TikTok. Hal ini dibuktikan pada hasil nilai evaluasi masing-masing perubahan parameter k. Akurasi tertinggi yang dicapai pada metode *K-Nearest Neighbor* sebesar 77.04% dengan parameter  $k = 11$  sedangkan pada metode *K-Nearest Neighbor* dengan Chi Square diperoleh sebesar 86.22%. Dengan parameter  $k=9$  yang mampu menghasilkan jumlah prediksi benar sebanyak 169 dari 196 data uji dengan hasil klasifikasi positif sebanyak 102 ulasan dan negatif sebanyak 94 ulasan.

#### Daftar Pustaka

- [1] C. M. Annur, "Indonesia Sabet Posisi Kedua Sebagai Negara Pengguna Tiktok Terbanyak Di Dunia Pada Awal 2023," *Databoks.Katadata.Co.Id*, 2023. <https://Databoks.Katadata.Co.Id/Datapublish/2023/02/27/Indonesia-Sabet-Posisi-Kedua-Sebagai-Negara-Pengguna-Tiktok-Terbanyak-Di-Dunia-Pada-Awal-2023> (Accessed Jun. 11, 2023).
- [2] S. Fide, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi," Vol. 10, No. 3, Pp. 346–358, 2021, [Online]. Available: <https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/>
- [3] J. Alfiah Zulqornain and P. Pandu Adikara, "Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes Dan Categorical Proportional Difference (Cpd)," 2021. [Online]. Available: <http://J-Ptiik.Ub.Ac.Id>
- [4] S. Khairunnisa, A. Adiwijaya, And S. Al Faraby, "Pengaruh Text Preprocessing Terhadap Analisis Sentimen Komentar Masyarakat Pada Media Sosial Twitter (Studi Kasus Pandemi Covid-19)," *Jurnal Media Informatika Budidarma*, Vol. 5, No. 2, P. 406, Apr. 2021, Doi: 10.30865/Mib.V5i2.2835.
- [5] R. Ahuja, A. Chug, S. Kohli, S. Gupta, And P. Ahuja, "The Impact of Features Extraction On The Sentiment Analysis," In *Procedia Computer Science*, Elsevier B.V., 2019, Pp. 341–348. Doi: 10.1016/J.Procs.2019.05.008.
- [6] D. Syafira, "Analisis Sentimen Ulasan Produk Kecantikan Menggunakan Metode Bm25 Dan Improved K-Nearest Neighbor Dengan Seleksi Fitur Chi-Square," 2020. [Online]. Available: <http://J-Ptiik.Ub.Ac.Id>
- [7] D. Musfiroh *Et AL.*, "Malcom: Indonesian Journal Of Machine Learning And Computer Science Sentiment Analysis Of Online Lectures In Indonesia From Twitter Dataset Using Inset Lexicon

- Analisis Sentimen Terhadap Perkuliahan Daring Di Indonesia Dari Twitter Dataset Menggunakan Inset Lexicon,” Vol. 1, Pp. 24–33, 2021.
- [8] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Vol. 12. Springer Science & Business Media, 2011.
- [9] P. Syahara, “Identifikasi Huruf Hijaiyah Tulisan Tangan Menggunakan Jaringan Syaraf Tiruan Backpropagation,” Smik Akakom Yogyakarta, 2018.

Halaman ini sengaja dibiarkan kosong