

# IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI DIGITAL KORLANTAS POLRI

Egi Puji Sutrisno<sup>1\*</sup>, Safrina Amini<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1\*</sup>[puijeg6@gmail.com](mailto:puijeg6@gmail.com) <sup>2</sup>[safrina.amini@budiluhur.ac.id](mailto:safrina.amini@budiluhur.ac.id)  
(\* : corresponding author)

**Abstrak**-Digital Korlantas POLRI adalah aplikasi untuk memperpanjang SIM secara online yang diluncurkan Korlantas POLRI untuk membantu masyarakat dalam memperpanjang SIM. Aplikasi Digital Korlantas POLRI sudah diunduh sebanyak kurang lebih 1 juta download dengan 87,1 ribu ulasan dan 3,8 rating. Banyaknya ulasan pengguna aplikasi sehingga memerlukan waktu untuk mengetahui secara keseluruhan ulasan pengguna. Dibutuhkan suatu sistem yang dapat mengklasifikasikan sentimen seseorang secara otomatis ke dalam kelas positif atau negatif. Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma K-Nearest Neighbor untuk mengklasifikasikan ulasan pengguna aplikasi. Metode yang digunakan adalah analisis sentiment, preprocessing data, pembobotan TF-IDF dan algoritma K-NN. Dataset yang digunakan pada penelitian ini sebanyak 600 data ulasan pengguna yang dibagi menjadi 70% data latih dan 30% data uji. Hasil pengujian didapati nilai akurasi tertinggi dengan nilai  $k=9$  sebesar 93.88%, nilai presisi tertinggi dengan nilai  $k=5$  sebesar 96.90% dan recall tertinggi dengan nilai  $k=9$  sebesar 93.26%. Berdasarkan dari 180 ulasan yang terdapat pada data uji, ulasan masyarakat terhadap aplikasi cenderung positif dengan 104 ulasan positif dan 76 ulasan negatif.

**Kata Kunci:** Analisis Sentimen, *K-Nearest Neighbor*, Digital Korlantas Polri, Klasifikasi, K-NN

## IMPLEMENTATION OF K-NEAREST NEIGHBOR ALGORITHM IN SENTIMENT ANALYSIS OF USER REVIEWS FOR DIGITAL KORLANTAS POLRI APPLICATION

**Abstract**-The Digital Korlantas POLRI is an application for extending driver's licenses (SIM) online, launched by the Indonesian National Police Traffic Corps (Korlantas POLRI) to assist the public in renewing their SIM. The Digital Korlantas POLRI application has been downloaded approximately one million times, with 87.1 thousand reviews and a 3.8 rating. Due to the large number of user reviews, it takes time to analyze all of them comprehensively. Therefore, a system is needed to automatically classify a person's sentiment into positive or negative classes. The aim of this research is to implement the K-Nearest Neighbor algorithm to classify user reviews of the application. The method involves sentiment analysis, data preprocessing, TF-IDF weighting, and the K-NN algorithm. The dataset used in this research consists of 600 user review data, divided into 70% training data and 30% test data. The results of the testing show the highest accuracy of 93.88% with  $k=9$ , the highest precision of 96.90% with  $k=5$ , and the highest recall of 93.26% with  $k=9$ . Based on the 180 reviews in the test data, it is found that the public's feedback towards the application is generally positive, with 104 positive reviews and 76 negative reviews.

**Keywords:** Sentiment Analysis, *K-Nearest Neighbor*, Digital Korlantas Polri, Classification, K-NN

## 1. PENDAHULUAN

Pesatnya perkembangan teknologi membawa Indonesia ke era serba digital. Di era yang serba digital seperti ini, secara umum manusia telah memiliki gaya hidup baru yang ketergantungan pada perangkat elektronik. Oleh karena itu, kemungkinan Indonesia akan meningkatkan upaya dalam melaksanakan berbagai program dan mempercepat transformasi digital, dengan mengintegrasikannya sebagai program nasional yang melibatkan hampir seluruh sektor industri[1].

Pada 13 April 2021 Korlantas Polri meluncurkan sebuah pelayanan Surat Izin Mengemudi (SIM) online melalui program (SINAR) atau SIM Presisi Nasional[2]. Dengan hadirnya layanan SIM Presisi Nasional (SINAR) melalui aplikasi Digital Korlantas POLRI, membuat proses melakukan perpanjangan SIM lebih mudah untuk masyarakat karena dilakukan secara *online*. Menurut informasi yang diperoleh dari situs web [Digitalkorlantas.id](https://digitalkorlantas.id),

pada aplikasi tersebut juga terdapat beberapa fitur lain yang dapat digunakan. Selain SINAR mereka juga menawarkan layanan SIGNAL (Samsat Digital Nasional) layanannya berupa pembayaran Pajak Kendaraan Bermotor (PKB)[3].

Korlantas Polri sangat terbuka terhadap segala masukan dari masyarakat terhadap pelayanan yang diberikan. Salah satunya yaitu terkait aplikasi Digital Korlantas POLRI sebagai aplikasi pelayanan perpanjangan SIM, masyarakat dapat memberikan masukan terkait aplikasi tersebut di kolom ulasan dari platform Google Play Store[4]. Saat ini pada *platform Google Play Store* aplikasi Digital Korlantas POLRI sudah diunduh sebanyak kurang lebih 1 juta download dengan 87,1 ribu ulasan dan 3,8 rating. Berbagai ulasan positif dan negatif serta rating dari aplikasi Digital Korlantas POLRI menunjukkan bahwa pelayanan yang diberikan masih belum sepenuhnya memenuhi harapan pengguna aplikasi tersebut. Karena banyaknya sentimen dari para pengguna aplikasi maka diperlukan waktu yang banyak untuk mengetahui suatu sentimen secara lengkap apakah termasuk sentimen positif atau negatif. Oleh sebab itu, dibutuhkan suatu sistem yang dapat mengklasifikasikan sentimen seseorang secara otomatis ke dalam kelas positif atau negatif.

Analisis sentimen adalah proses menggunakan Bahasa alami untuk memahami pendapat orang tentang produk atau topik tertentu, analisis sentimen untuk pengumpulan dan penelitian pendapat yang diungkapkan dalam blog, postingan atau komentar media sosial[5]. Analisis sentimen dapat diterapkan pada berbagai opini, seperti opini publik, film, politik, ujaran kebencian dan lainnya di media sosial. Dari keterangan tersebut, dapat disimpulkan bahwa analisis sentimen dapat digunakan untuk menganalisis informasi yang ada di media sosial, seperti pesan atau ulasan terkait instansi atau perusahaan[6]. Di dalam informasi tersebut terdiri dari dua kelas yang mencakup ulasan positif dan negatif.

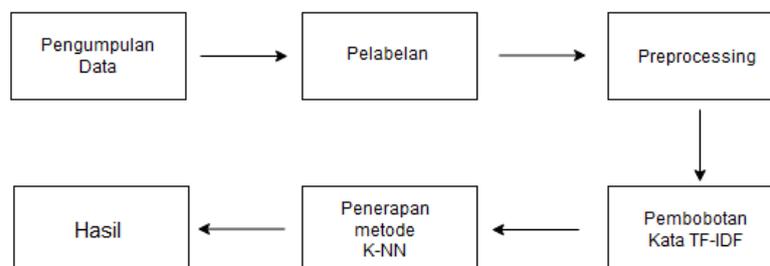
Beberapa penelitian sebelumnya tentang analisis sentimen diantaranya adalah penelitian yang membahas tentang Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor didapati nilai akurasi terbesar 67,2% dengan  $k=5$ , nilai presisi tertinggi sebesar 56,94% saat  $k=5$  dan recall sebesar 78,24% saat  $k=15$  tanpa konversi ke kata baku [7]. Penelitian lainnya yang sudah dilakukan tentang Analisis Sentimen Ulasan Aplikasi WeTV Untuk Peningkatan Layanan Menggunakan Metode K-Nearest Neighbor didapati dengan nilai  $k=3$  pada model dataset 90% data latihan dan 10% data uji memperoleh akurasi sebesar 0,70% precision 0,76% recall dan 0,72% f1-score dengan penghitungan euclidean [8].

Penelitian sebelumnya menunjukkan tanpan konversi kata baku dan penggunaan penghitungan menggunakan euclidean untuk algoritma *K-Nearest Neighbor* (KNN), peneliti tertarik untuk menggunakan konversi kata baku dan penghitungan *cosine similarity* untuk algoritma *K-Nearest Neighbor* (KNN) dalam menganalisis sentimen. Sistem analisis sentimen ini akan dibuat dengan aplikasi berbasis web dan metode yang akan digunakan dalam penelitian ini adalah *web scraping* untuk pengumpulan data dan digunakan algoritma *K-Nearest Neighbor* (KNN) yang dikombinasikan dengan metode *Term Frequency –Inverse Document Frequency* (TF-IDF) untuk klasifikasi dan pembobotan katanya.

## 2. METODE PENELITIAN

### 2.1 Alur Keseluruhan Sistem

Berikut akan ditampilkan tahapan-tahapan yang akan dilakukan secara keseluruhan.



Gambar 1. Alur Keseluruhan Sistem

Gambar 1 menunjukkan tahapan-tahapan yang akan dilakukan pada penelitian ini. Tahapan yang paling pertama yaitu pengumpulan data melalui proses *scraping* pada halaman Digital Korlanta POLRI yang ada di

Google Play Store. Kemudian data yang terkumpul akan dilakukan pelabelan secara manual menjadi dua kelas yaitu, kelas positif dan kelas negatif. Setelah dilakukan pelabelan, data akan di preprocessing untuk mengurangi kata noise dan menghasilkan data yang bersih. Setelah itu data akan dilakukan pembobotan kata menggunakan TF-IDF, setelah itu akan dilakukan penerapan metode KNN

## 2.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan teknik *scraping* menggunakan bahasa pemrograman python menggunakan pustaka `google_play_scraper`. Data diambil dari halaman aplikasi Digital Korlantas POLRI yang ada pada platform Google Play Store, setelah itu data akan disimpan ke dalam sebuah file berformat csv. Selanjutnya data akan dibagi menjadi data latih dan data uji, lalu akan diberi label dan data latih akan diimport ke dalam database. Data yang diambil dari rentang waktu 5 April 2023 hingga 27 April 2023 dan 12 Mei 2023 hingga 31 Mei 2023, yang menghasilkan 600 total data ulasan.

## 2.3 Pelabelan Data

Setelah mengumpulkan data pada tahap pengumpulan data, pada tahap pelabelan ini, setiap data diberi label kedalam salah satu kelas berupa positif atau negatif. Akan diberi label positif bila pengguna merasakan manfaat dan terbantu oleh aplikasi. Akan diberi label negatif bila pengguna kurang terbantu atau tidak mendapatkan manfaat dari aplikasi tersebut, serta mengalami kesusahan dan belum merasakan fitur-fitur yang ada dalam aplikasi.

## 2.4 Preprocessing

*Preprocessing* adalah tahapan awal dalam mempersiapkan dan membersihkan teks dari kata-kata yang tidak berguna. Pada tahapan ini terdapat beberapa langkah untuk membersihkan teks dari noise atau informasi yang tidak relevan, untuk memastikan konsistensi dan kualitas data yang akan digunakan dalam analisis lebih lanjut. Beberapa tahap preprocessing yang akan digunakan untuk penelitian ini, yaitu *case folding*, *symbol removal*, normalisasi, *stopword removal* dan *stemming*.

### a. Case Folding

*Case folding* merupakan proses merubah huruf kapital menjadi huruf non kapital agar teks mempunyai struktur yang sama dan mudah untuk di proses.

### b. Symbol Removal

*Symbol Removal* merupakan proses untuk penghapusan simbol-simbol, tanda baca (titik (.), koma(,), tanda Tanya(?) dan lain-lain), angka numerik (0-9) dan karakterlainnya (@, #, \$, dan lain-lain).

### c. Normalisasi

Normalisasi merupakan proses yang bertujuan untuk menormalisasikan dari setiap kata yang mengandung kata gaul, kata tidak baku, atau singkatan kata tertentu menjadi kata baku.

### d. Stopword Removal

*Stopword removal* merupakan proses menghapus kata-kata yang tidak memiliki makna. Tujuan dari tahap ini adalah mengurangi jumlah kata yang akan digunakan untuk model latih agar menghasilkan dimensi yang lebih efisien.

### e. Stemming

*Stemming* merupakan proses pengolahan kata-kata untuk mendapatkan kata dasar dari sebuah kata yang telah mengalami imbuhan dengan aturan-aturan tertentu. Proses stemming ini dibantu library sastrawi.

## 2.5 Pembobotan Kata

TF-IDF adalah suatu metode untuk menghitung bobot setiap kata yang paling umum digunakan dalam information retrieval. Metode Term Frequency - Inverse Document Frequency adalah sebuah pendekatan yang digunakan untuk memberikan bobot antara sebuah kata (term) dengan dokumen. TFIDF merupakan suatu ukuran statistik yang digunakan untuk mengukur seberapa penting sebuah kata dalam dokumen atau di dalam sekelompok kata[9]. Pada metode TF-IDF digunakan rumus seperti pada persamaan (1), (2) dan (3).

$$TF(t) = (f(d,t))/(n(d,t)) \quad (1)$$

$$IDF(t) = (\log(Nd))/(df(t)) \quad (2)$$

$$TF-IDF = TF(t) \times IDF(t) \quad (3)$$

Keterangan:

TF (t) : Term frekuensi pada term t

- $f(d,t)$  : Frekuensi kemunculan term  $t$  pada dokumen  $d$   
 $n(d,t)$  : Jumlah seluruh term kata pada suatu dokumen  
 $IDF(t)$  : *Inverse* document frequency pada term  $t$   
 $Nd$  : Jumlah keseluruhan dokumen  
 $df(t)$  : Jumlah dokumen yang mengandung term  $t$   
 $TF-IDF$  : *term frequency* dikali *inverse document frequency*

## 2.6 K-Nearest Neighbor

*K-Nearest Neighbor* adalah salah satu metode *supervised learning* yang melibatkan proses pembelajaran dari data training. Algoritme K-NN sangat umum digunakan dalam melakukan klasifikasi objek karena pengaplikasiannya sangat sederhana, efektif dan efisien [10]. Dalam algoritma ini, klasifikasi dilakukan berdasarkan jarak kemiripan antara fitur-fitur pada data training yang ada menggunakan rumus cosine similarity, lalu sistem akan mengurutkan nilai jarak dari yang terendah sampai yang tertinggi. Setelah diketahui nilai jarak dari yang terendah sampai yang tertinggi, akan diambil sebesar  $k$  data nilai tertingginya. Dari  $k$  data tersebut akan dianalisis untuk melihat sentimen mana yang paling sering muncul. Kelas sentimen yang paling banyak muncul merupakan kelas atau nilai dari sentimen data yang sedang dihitung. Berikut rumus untuk menghitung persamaan cosine similarity pada persamaan (4).

$$\text{similarity}(a, b) = \frac{A \cdot B}{|A| \cdot |B|} = \frac{A_i \cdot B_i}{\sqrt{\sum_{i=1}^n (A)^2} \times \sqrt{\sum_{i=1}^n (B)^2}} \quad (4)$$

Keterangan :

- $A$  = Vektor A yang akan dibandingkan  
 $B$  = Vektor B yang akan dibandingkan  
 $A \cdot B$  = *dot product* antara vektor A dan vektor B  
 $|A|$  = panjang vektor A  
 $|B|$  = panjang vektor B  
 $|A| \cdot |B|$  = *cross product* antara panjang vektor A dan panjang vektor B

## 2.7 Pengujian

Pengujian dilakukan untuk mengetahui nilai atau tingkat akurasi, presisi dan recall dari klasifikasi menggunakan KNN. Pada penelitian ini, dilakukan dengan cara membandingkan data hasil prediksi dengan data aktual dengan bantuan *confusion matrix*.

**Tabel 1.** *Confusion Matrix*

<i>Actual Value</i>	<i>Predicted Values</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Dengan memanfaatkan confusion matrix dapat diperoleh nilai akurasi, presisi dan recall. Berikut rumus untuk menghitung akurasi, presisi dan recall yang dapat dilihat pada persamaan (5), (6) dan (7).

$$\text{akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

$$\text{presisi} = \frac{TP}{(TP + FP)} \quad (6)$$

$$\text{recall} = \frac{(TP)}{(TP + FN)} \quad (7)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Dataset

Penelitian ini menggunakan dataset yang didapat dari halaman Google Play Store. Data tersebut diambil menggunakan proses scraping dengan pustaka `google_play_scraper`. Jumlah total data yang akan digunakan untuk dataset sebanyak 600 data. Contoh data yang digunakan dapat ditemukan pada tabel 2.

Tabel 2. Contoh *Dataset*

<i>Username</i>	<i>Content</i>	<i>Tanggal</i>
Ahmad khoirur rozak Rozak	Aplikasinya <i>error</i> gak bisa akese saat ini	4/26/2023

#### 3.2 Pelabelan Data

Dalam tahap pelabelan, data yang sudah dikumpulkan akan diberi label secara manual. Akan diberi label positif bila pengguna merasakan manfaat dan terbantu oleh aplikasi. Akan diberi label negatif bila pengguna kurang terbantu atau tidak mendapatkan manfaat dari aplikasi tersebut, serta mengalami kesusahan dan belum merasakan fitur-fitur yang ada dalam aplikasi. Berikut contoh pelabelan data yang dapat dilihat pada tabel 3.

Tabel 3. *Dataset* yang sudah diberi label

<i>Username</i>	<i>Content</i>	<i>Label</i>
Ahmad khoirur rozak Rozak	Aplikasinya <i>error</i> gak bisa akese saat ini	negatif

#### 3.3 Preprocessing

Pada tahap *preprocessing*, dilakukan proses transformasi pada data yang diperoleh, dengan tujuan agar menghasilkan data yang telah bersih sehingga dapat diproses oleh sistem. Data yang sudah di-*preprocessing* akan digunakan untuk mengembangkan model kata yang akan digunakan dalam proses klasifikasi. Tahapan-tahapan *preprocessing* dapat dilihat pada tabel 4.

Tabel 4. Tahapan *preprocessing*

<b>Tahapan</b>	<b>Data Awal</b>	<b>Hasil Proses</b>
<i>Case Folding</i>	Aplikasinya <i>error</i> gak bisa akese saat ini	aplikasinya <i>error</i> gak bisa akese saat ini.
<i>Symbol Removal</i>	aplikasinya <i>error</i> gak bisa akese saat ini.	aplikasinya <i>error</i> gak bisa akese saat ini
Normalisasi	aplikasinya <i>error</i> gak bisa akese saat ini	aplikasinya <i>error</i> tidak bisa akses saat ini
<i>Stopword Removal</i>	aplikasinya <i>error</i> tidak bisa akses saat ini	aplikasinya <i>error</i> akses
<i>Stemming</i>	aplikasi <i>error</i> akses	aplikasi <i>error</i> akses

#### 3.4 Pembobotan Kata

Setelah melalui tahapan *preprocessing*, langkah selanjutnya yaitu melakukan pembobotan kata menggunakan metode TF-IDF. Pembobotan kata merupakan proses mengubah kata-kata menjadi angka-angka yang mempresentasikan bobot atau nilai pentingnya suatu dokumen dalam sekumpulan data. contoh data yang akan dilakukan pembobotan TF-IDF dapat dilihat pada tabel 5.

Tabel 5. Contoh data latihan

<b>No</b>	<b>Ulasan</b>	<b>Label</b>
D1	sangat bantu puas layan	positif
D2	akses buruk sekali	negatif

Langkah pertama yaitu menghitung *Term Frequency* (TF). Dalam perhitungan ini, dilakukan pencarian kemunculan setiap kata dalam dokumen yang kemudian dibagi dengan jumlah kata dalam dokumen tersebut, rumus penghitungan (TF) dapat dilihat dipersamaan (1). Lalu dilanjutkan mencari *document frequency* (DF) pada setiap dokumen, selanjutnya adalah mencari *inverse document frequency* (IDF) menggunakan logaritma rumus menghitung IDF dapat dilihat dipersamaan (2). Setelah itu nilai IDF yang telah didapat akan digunakan untuk menghitung bobot kata (TF-IDF) dengan mengalikan nilai TF dengan Nilai IDF. Pada tabel 6 adalah tabel penghitungan TF, DF, IDF dan TF-IDF.

**Tabel 6.** Pembobotan kata TF-IDF data latih

Term	TF D1	TF D2	DF	IDF	TF-IDF D1	TF-IDF D1
akses	0	0.33	1	0.30103	0	0.1003
bantu	0.25	0	1	0.30103	0.0753	0
buruk	0	0.33	1	0.30103	0	0.1003
layan	0.25	0	1	0.30103	0.0753	0
puas	0.25	0	1	0.30103	0.0753	0
sangat	0.25	0	1	0.30103	0.0753	0
sekali	0	0.33	1	0.30103	0	0.1003

### 3.5 Implementasi Metode K-Nearest Neighbor

Pada tahap ini, akan mengklasifikasikan menggunakan K-Nearest Neighbor (KNN) setelah melakukan pembobotan TF-IDF. Tahapan ini dilakukan untuk memprediksi label dari data yang akan diuji. Berikut ini adalah contoh data uji yang sudah melewati preprocessing yang dapat ditemukan pada tabel 7.

**Tabel 7.** Data uji

Data Uji
mantap sangat bantu sangat

Lalu data uji akan dilakukan pembobotan kata atau merubah kata-kata ke representasi vektor angka untuk dilakukan penghitungan lebih lanjut, pada tabel 8 merupakan tabel hasil TF-IDF data uji

**Tabel 8.** Pembobotan kata TF-IDF data uji

Term	TF-IDF Data Uji
akses	0
bantu	0.1003
buruk	0
layan	0
puas	0
sangat	0.2007
sekali	0

Setelah mendapatkan vector angka data uji, selanjutnya akan dilakukan penghitungan kemiripan Antara data uji dengan data latih menggunakan rumus cosine similarity yang dapat dilihat pada persamaan (4). Dalam rumus

cosine similarity, langkah pertama yaitu mencari nilai dot product Antara data uji dengan data latih yang dapat dilihat pada tabel 9.

**Tabel 9.** Menghitung *dot product*

	Data uji • D1	Data uji • D2
	0x0	0x0.1003
	0.1003x0.0753	0.1003x0
	0x0	0x0.1003
	0x0. 0753	0x0
	0x0. 0753	0x
	0.2007 x0. 0753	0.2007x0
	0x0	0x0.1003
Hasil <i>dot product</i>	0.0226653	0

Selanjutnya setelah mendapatkan nilai dot product, akan dilakukan penghitungan untuk mengetahui nilai panjang vektor data uji dan data latih yang dapat ditemukan pada tabel 10.

**Tabel 10.** Menghitung panjang vektor

	Vektor D1	Vektor D2	Vektor Data Uji
	0 <sup>2</sup>	0.1003 <sup>2</sup>	0 <sup>2</sup>
	0.0753 <sup>2</sup>	0 <sup>2</sup>	0.1003 <sup>2</sup>
	0 <sup>2</sup>	0.1003 <sup>2</sup>	0 <sup>2</sup>
	0.0753 <sup>2</sup>	0 <sup>2</sup>	0 <sup>2</sup>
	0.0753 <sup>2</sup>	0 <sup>2</sup>	0 <sup>2</sup>
	0.0753 <sup>2</sup>	0 <sup>2</sup>	0.2007 <sup>2</sup>
	0 <sup>2</sup>	0.1003 <sup>2</sup>	0 <sup>2</sup>
$\sum n^2$	0.02268036	0.03018027	0.05034508
$\sqrt{\sum n^2}$	0.1506	0.1737246959991584	0.2243770933049985

Setelah mengetahui nilai panjang vektornya, untuk menentukan nilai *cosine similarity* dilakukan pembagian antara hasil dot product dibagi dengan panjang vektor data uji dikali dengan panjang vektor data latih. Maka didapatkan nilai cosine similarity, nilai cosine similarity dapat dilihat di tabel 11.

**Tabel 11.** Hasil penghitungan *cosine similarity*

Similarity (Data Uji, D1)	Similarity (Data Uji, D1)
0.6707	0

Dari hasil penghitungan *cosine similarity* akan diurutkan dari nilai terkecil hingga terbesar dan akan diambil nilai terbesar sebanyak nilai k.

- D2 Negatif (0)
- D1 Positif (0.6707)

Jika k=1 maka akan diambil 1 nilai tertinggi, dari nilai yang diambil akan dilihat kelas mana yang paling banyak muncul dan akan diklasifikasikan berdasarkan kelas yang paling banyak muncul. Pada contoh ini, data uji akan diklasifikasikan ke label positif.

### 3.6 Pengujian

Pada tahap pengujian, menggunakan data sebanyak 420 data latih yang berisi 210 ulasan positif dan 210 ulasan negatif dan menggunakan 180 data uji. Tujuan dari pengujian ini untuk mengevaluasi sejauh mana system dapat mengklasifikasikan suatu ulasan. Pengujian akan menggunakan tujuh nilai k yaitu k=3, k=5, k=7, k=9, k=11, k=13, k=15 untuk mengetahui seberapa besar akurasi yang didapat dari ketujuh nilai k tersebut. Keseluruhan data hasil prediksi data uji kemudian direpresentasikan ke dalam *confusion matrix*, representasi *confusion matrix* untuk k=3 dapat dilihat pada tabel 12.

**Tabel 12.** *Confusion Matrix* k=3

Actual Value	Predicted Values	
	Positive	Negative
Positive	95	9
Negative	6	70

Berdasarkan tabel 12 dari *confusion matrix* dengan  $k=3$ , hasil nilai akurasi, presisi, dan recall dapat dihitung menggunakan rumus-rumus yang dijelaskan dalam persamaan (5), (6), dan (7) seperti yang tertera pada tabel 14.

**Tabel 13.** Penghitungan *Confusion Matrix*  $k=3$

Pengujian		
Akurasi	$= \frac{95 + 70}{95 + 70 + 6 + 9}$	= 0.9166
Presisi	$= \frac{95}{95 + 6}$	= 0.9405
Recall	$= \frac{95}{95 + 9}$	= 0.9134

Pengujian di atas dilakukan berulang-ulang dengan variasi nilai  $K$  yang berbeda-beda. Sehingga dapat diketahui hasil pengujian secara keseluruhan adalah seperti tabel 14.

**Tabel 14.** Hasil Pengujian

	K=3	K=5	K=7	K=9	K=11	K=13	K=15
Akurasi	0.9166	0.9277	0.9277	0.9388	0.9222	0.9277	0.9111
Presisi	0.9405	0.9690	0.9690	0.9603	0.95	0.9595	0.9583
Recall	0.9134	0.9038	0.9038	0.9326	0.9134	0.9134	0.8846

Berdasarkan tabel hasil pengujian dapat diketahui bahwa nilai pengujian dengan akurasi tertinggi didapatkan pada nilai  $k=9$  sebesar 93.88%, presisi tertinggi didapatkan pada nilai  $k=5$  sebesar 96.90%, dan recall tertinggi sebesar 93.26% dengan  $k=9$ . Sementara itu hasil sentimen terhadap 180 data uji menunjukkan ulasan pengguna aplikasi Digital Korlantas POLRI cenderung positif sebanyak 104 ulasan positif dan 76 ulasan negatif. Pada penelitian sebelumnya dengan judul Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor didapatkan nilai akurasi terbesar 67,2% dengan  $k=5$ , nilai presisi tertinggi sebesar 56,94% saat  $k=5$  dan recall sebesar 78,24% saat  $k=15$  tanpa konversi ke kata baku [7]. Penggunaan konversi bahasa gaul ke bentuk baku dapat meningkatkan nilai akurasi yang didapat.

#### 4. KESIMPULAN

Hasil pengujian menunjukkan bahwa sistem mampu secara otomatis mengklasifikasikan ulasan ke dalam kelas positif atau negatif menggunakan metode KNN. Tahapan utama dalam penelitian ini melibatkan proses scraping, pelabelan, preprocessing, pembobotan TF-IDF, dan klasifikasi K-Nearest Neighbor. Kualitas tahap preprocessing menjadi faktor penentu dalam mencapai hasil yang optimal. Penambahan normalisasi dari bahasa gaul ke bentuk baku terbukti efektif dalam membersihkan noise yang terdapat dalam data. Penggunaan pembobotan kata TF-IDF dan algoritma K-Nearest Neighbor dalam analisis sentimen berjalan dengan baik dan menghasilkan hasil yang memuaskan. Dari 180 sentimen yang dianalisis dalam data uji, mayoritas masyarakat memberikan sentimen positif (104 sentimen positif) terhadap aplikasi, sementara 76 sentimen cenderung negatif. Tingkat akurasi tertinggi tercapai saat  $k$  bernilai 9, mencapai 93.88%. Tingkat presisi tertinggi tercapai pada  $k$  bernilai 5, mencapai 96.90%, sementara recall tertinggi, yaitu 93.26%, terjadi pada  $k$  bernilai 9.

Sebagai saran untuk pengembangan lebih lanjut, disarankan untuk meningkatkan jumlah data latih agar variasi data lebih baik dan menghasilkan akurasi yang lebih tinggi. Penambahan daftar kata slang dan stopword dalam tahap preprocessing juga dianjurkan untuk membersihkan noise dengan lebih efektif. Selain itu, penelitian berikutnya dapat mempertimbangkan untuk mengajukan bahasa-bahasa lain selain bahasa Indonesia dalam analisis sentimen.

## DAFTAR PUSTAKA

- [1] W. Setiawan, “Era Digital dan Tantangannya. Seminar Nasional Pendidikan,” *Semin. Nas. Pendidik.*, pp. 1–9, 2017.
- [2] (2021) Kompas. [Online]. Available: <https://nasional.kompas.com/read/2021/04/13/18003571/korlantas-polri-luncurkan-aplikasi-sinar-perpanjang-sim-cukup-lewat>
- [3] (2023) Digitalkorlantas Polri. [Online]. Available: <https://www.digitalkorlantas.id/>
- [4] E. R. Kaburuan and N. R. Setiawan, “Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, Mar. 2023.
- [5] (2023) Qeios website. [Online]. Available: <https://www.qeios.com/read/YF9X04>
- [6] S. Bhatia, M. Sharma, and K. K. Bhatia, “Sentiment Analysis and Mining of Opinions,” *Stud. Big Data*, vol. 30, no. May, pp. 503–523, 2018.
- [7] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018.
- [8] N. Faridhotun, E. Haerani, and R. M. Candra, “Analisis Sentimen Ulasan Aplikasi WeTV Untuk Peningkatan Layanan Menggunakan Metode K-Nearst Neighbor,” *J. Inf. Syst. Res.*, vol. 4, no. 3, pp. 855–864, Apr. 2023.
- [9] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018.
- [10] N. N. I. W. Astuti, I. M. A. D. Suarjaya, and I. M. S. Raharja, “Analisis Sentimen Masyarakat Terhadap Pelaksanaan Layanan Kesehatan Selama Masa Pandemi Di Indonesia Menggunakan Teknologi Big Data,” *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 2, p. 1245, 2022.