

## ANALISIS SENTIMEN MASYARAKAT TERHADAP TILANG ELEKTRONIK PADA TWITTER DENGAN ALGORITMA K-NEAREST NEIGHBOR

Hendriansyah<sup>1\*</sup>, Wahyu Pramusinto<sup>2</sup>

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

Email: <sup>1\*</sup>1911501169@student.budiluhur.ac.id, <sup>2</sup> wahyu.pramusinto@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak**-Sistem transportasi merupakan suatu hal yang penting bagi suatu kota, terutama di kota-kota besar yang padat aktivitas dan padat penduduk. Selain itu, sistem transportasi sangat penting dalam menentukan efisiensi suatu kota. Banyak kasus pelanggaran lalu lintas di jalan raya yang cenderung berujung pada kecelakaan dan menambah kemacetan lalu lintas. Pelanggaran lalu lintas sebagian besar disebabkan oleh ketidakpatuhan terhadap peraturan yang terkait dengan jalur, rambu lalu lintas dan lampu lalu lintas, termasuk peraturan larangan berhenti, parkir di tempat yang tidak sah, lampu merah, mengemudi kendaraan tanpa izin dan perlengkapan kendaraan dan pelanggaran lainnya. Polisi telah menetapkan peraturan tentang tilang elektronik yang akan membantu menangani pelanggaran lalu lintas. Tujuan dari penelitian ini adalah Bagaimana presentasi pandangan (sentimen) masyarakat Indonesia terhadap tilang elektronik pada periode 20 maret hingga 29 mei 2023 Adapun masalah penelitian ini bagaimana nilai akurasi yang diperoleh algoritma *K-Nearest Neighbor* dalam melakukan analisis sentiment, Bagaimana cara menganalisis sentimen berdasarkan pendapat masyarakat Indonesia melalui media sosial twitter. *K-Nearest Neighbor* (K-NN) adalah salah satu algoritma yang digunakan dalam masalah klasifikasi. Prinsip operasi K-NN adalah mencari jarak terpendek antara data yang akan dievaluasi dengan tetangga terdekat. Metode yang digunakan adalah mengekstraksi fitur CountVectorizer dan algoritma klasifikasi K-Nearest Neighbor. Penelitian ini menggunakan library Tweet-harvest dengan menggunakan Google Colabs dan bahasa pemrograman Python. Kumpulan data yang digunakan diambil dari jejaring sosial Twitter berupa (tweet) atau kicauan dalam bahasa Indonesia didapatkan melalui fungsi pencarian dengan kata kunci "tilang elektronik". Hasil analisis berdasarkan 509 tweet menunjukkan 82,71 komentar positif dan 17,29% komentar negatif antara tanggal 20 Maret dan 29 Mei 2023, sedangkan hasil Uji agregat Hasil terbaik diperoleh dengan menggunakan nilai K = 5, dengan akurasi 68 85% akurasi, 77% ingat.

**Kata Kunci:** *counvectorizer, k-nearest neighbor, twitter, tilang elektronik*

### ANALYSIS OF PUBLIC SENTIMENT ON ELECTRONIC TILANG ON TWITTER SOCIAL MEDIA USING COUNTVECTORIZER FEATURE EXTRACTION AND K-NEAREST NEIGHBOR ALGORITHM

**Abstract**-The transportation system is an important thing for a city, especially in big cities that are full of activities and densely populated. In addition, the transportation system is very important in determining the efficiency of a city. Many cases of traffic violations on the highway tend to lead to accidents and increase traffic jams. Traffic violations are mostly caused by non-compliance with regulations related to lanes, traffic signs and traffic lights, including regulations prohibiting stopping, parking in unauthorized places, red lights, driving vehicles without permission and vehicle equipment and other violations. The police have established regulations on electronic ticketing that will help deal with traffic violations. The purpose of this research is to find How is the presentation of the views (sentiment) of the Indonesian people towards electronic ticketing from 20 March to 29 May 2023 as for the problem of this research What is the accuracy value obtained by the *K-Nearest Neighbor* algorithm in conducting sentiment analysis, How do you analyze sentiment based on the opinions of the Indonesian people through social media Twitter, *K-Nearest Neighbor* (K-NN) is one of the algorithms used in classification problems. The principle of K-NN operation is to find the shortest distance between the data to be evaluated and the nearest neighbor. The method used is extracting the CountVectorizer feature and the *K-Nearest Neighbor* classification algorithm. This research uses the Tweet-harvest library using Google Colabs and the Python programming language. The data set used was taken from the Twitter social network in the form of (tweets) or tweets in Indonesian which were obtained through a search function with the keyword "electronic ticket". The results of the analysis based on 509 tweets show that positive sentiment is 82.71 and negative is 17.29% between March 20 and May 29 2023, while the test results The best results are obtained using a value of K = 5, with an accuracy of 68%, and precision 85% and recall is 77%.

**Keywords:** *twitter, electronic tickets, counvectorizer, k-nearest neighbor*

## 1. PENDAHULUAN

Sistem transportasi merupakan suatu hal yang penting bagi suatu kota, terutama di kota-kota besar yang padat aktivitas dan padat penduduk. Selain itu, sistem transportasi sangat penting dalam menentukan efisiensi suatu kota. Banyak kasus pelanggaran lalu lintas oleh pengguna jalan yang cenderung berujung pada kecelakaan dan menambah kemacetan lalu lintas. Pelanggaran lalu lintas mayoritas berupa pelanggaran di tingkat sinyal, rambu, dan lampu lalu lintas seperti larangan berhenti, parkir di tempat tertentu, menerobos lampu merah, tidak berdokumen, suku cadang kendaraan.

Pelanggaran tersebut terjadi pada saat puncak aktivitas masyarakat di jalan raya. Pertumbuhan jumlah kendaraan bermotor di Indonesia meningkat pesat seiring dengan bertambahnya jumlah kendaraan bermotor serta meningkatnya pelanggaran lalu lintas. Platform inilah yang mendorong kepolisian untuk menetapkan peraturan denda elektronik untuk membantu menangani pelanggaran lalu lintas dan pungutan liar yang belakangan ini menghambat perkembangan moda transportasi.

Pelanggaran lalu lintas tidak dapat diabaikan begitu saja karena sebagian besar kecelakaan lalu lintas disebabkan oleh faktor manusia pengguna jalan yang tidak menaati peraturan lalu lintas. Namun masih ada penyebab selain faktor manusia seperti ban kempes, rem kempes, jalan berlubang, macet, jumlah kendaraan yang melewati beberapa jalan utama, kondisi jalan, perkerasan jalan dan infrastruktur lainnya tidak sesuai. Mengimplementasikan rancangan model yang dibuat ke dalam aplikasi berbasis web. Melakukan analisis sentimen masyarakat terhadap pandangan masyarakat Indonesia mengenai sistem tilang elektronik pada media sosial Twitter. Menguji keakuratan algoritma *K-Nearest Neighbor* dalam melakukan analisis sentimen. Pemasalahan penelitian ini yaitu bagaimana nilai akurasi yang diperoleh algoritma *K-Nearest Neighbor* dalam melakukan analisis sentiment, bagaimana cara menganalisis sentimen berdasarkan pendapat masyarakat Indonesia melalui media sosial twitter.

Adapun manfaat dan harapan dalam penelitian ini yaitu manfaat dari penelitian adalah menganalisis pandangan masyarakat Indonesia berdasarkan *tweet* yang dipublikasikan melalui media sosial twitter. Sehingga dapat diperoleh gambaran sentimen terkait topik pandangan masyarakat Indonesia terhadap sistem tilang elektronik di Indonesia. Hasil penelitian ini diharapkan menjadi bahan evaluasi bagi kepolisian Republik Indonesia. Penelitian ini juga dilakukan untuk menguji kinerja dan nilai akurasi algoritma *K-Nearest Neighbor* untuk analisis sentimen. Karena pemilihan metode tersebut dapat memengaruhi hasil dari penelitian yang terfokus pada masalah yang diteliti, dan memudahkan dalam proses analisis data.

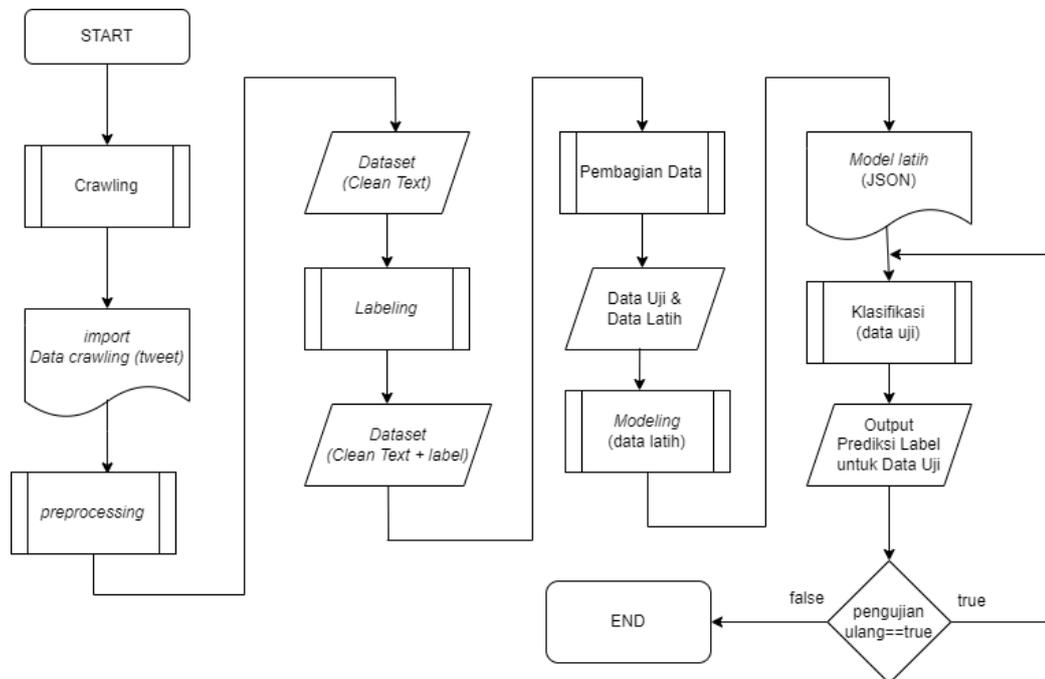
Beberapa penelitian sebelumnya tentang sentimen analisis diantaranya penelitian pertama membahas tentang Penerapan text mining pada aplikasi Tokopedia menggunakan algoritma *K-Nearest Neighbor* [1], mengklaim bahwa algoritma *K-Nearest Neighbor* menggunakan hingga 5.254 data terpisah menggunakan validasi silang *K-fold*, yang kemudian dihitung keakuratannya. dan parameter  $K=200$  sebesar 71,12%. Dalam penelitian lainnya Analisis sentimen new normal periode covid-19 di Indonesia [2], menyebutkan data tersebut mencakup 1000 tweet untuk sentimen positif, 811 dan 189 negatif. Pengklasifikasi *K-NN* dengan nilai  $K=1$  menghasilkan set pelatihan yang menggunakan akurasi 100%, 92,60% untuk validasi silang 10 kali lipat, dan 94,50% untuk distribusi persentil 80% percentage split. *KNN* juga menggunakan kajian berjudul analisis sentimen PT pelayanan PLN di Jakarta melalui Twitter [3], penelitian melakukan klasifikasi dengan algoritma *K-Nearest Neighbor* menggunakan jarak Euclidean dengan jarak antar tetangga memiliki total  $K=3$ . Dengan nilai akurasi 89,4 dengan hasil pemerinkatan pada bulan Februari 2020 di DKI Jakarta dari total 500.

Penelitian ini bertujuan untuk mengetahui penilaian masyarakat terhadap penerapan tilang elektronik pada maret hingga mei 2023. Metode yang digunakan adalah melakukan analisis sentimen melalui machine learning dengan fungsi kamus sentimen, dengan ekstraksi fitur menggunakan algoritma klasifikasi *CountVectorizer* dan *K-Nearest Neighbor*. Dataset yang digunakan adalah tweet dari jejaring sosial Twitter dengan keyword “e-ticket”. Pengumpulan kumpulan data dilakukan dari 20 Maret 2023 hingga 29 Mei 2023. Sebanyak 509 tweet dikumpulkan menggunakan kata kunci dan rentang tanggal yang disarankan.

## 2. METODE PENELITIAN

Ada pun beberapa proses yaitu *crawling*, *preprocessing labelling*, pembagian data, *modelling* dan pengujian. *Flowchart* Alur pencarian dapat dilihat pada Gambar 1. Pengumpulan data dilakukan dengan proses data mining untuk mendapatkan kumpulan data berupa tweet. Selain itu, semua tweet dibuat di Excel, lalu diimpor ke database untuk diproses sebelumnya. Hasil *preprocessing* menghasilkan kalimat yang lebih terstruktur (*clean text*) yang kemudian dieksekusi pada langkah selanjutnya. *Clean text* yang diperoleh dari *preprocessing* akan diproses pada tahap labeling untuk menentukan klasifikasi (label) positif atau negatif, setelah itu tweet yang diberi label akan dibagi menjadi dua (2) bagian, detailnya: data uji dan data latih.

Data pelatihan merupakan data yang digunakan sebagai pengetahuan untuk proses klasifikasi, pengembangan dilakukan dengan memodelkan dan membuat model pelatihan serta menggunakan data pelatihan yang tersedia. Sedangkan data uji adalah data yang disiapkan untuk menguji akurasi model pelatihan yang dihasilkan dari proses pemodelan, verifikasi akurasi dilakukan melalui prosedur pelatihan, uji, uji dan evaluasi. Setelah proses pengujian dan evaluasi selesai, hasil penerapan metode tersebut akan ditampilkan secara visual berupa persentase dan grafik.



Gambar 1. Flowchart Penelitian

## 2.1 Text Mining

Text mining adalah proses mendapatkan informasi berkualitas tinggi dari teks [4], text mining adalah teknik yang dapat digunakan untuk melakukan klasifikasi dokumen, clustering, ekstraksi informasi, analisis, dll. sentimen dan pengambilan informasi dimana text mining adalah varian dari data mining yang berupaya menemukan pola menarik dari sejumlah besar data tekstual [5].

## 2.2 Analisis Sentimen

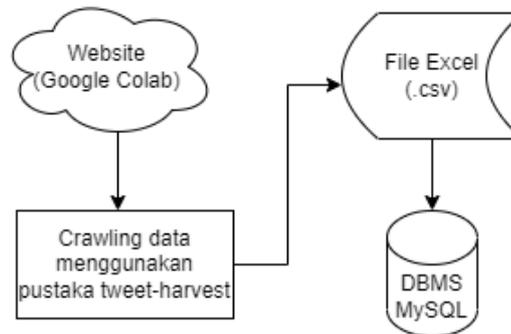
Analisis sentimen adalah metode yang digunakan untuk mengekstrak data opini, memahami dan mengolah data tekstual secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini [6].

## 2.3 Media Sosial

Media sosial adalah media sosial online, pengguna dapat dengan mudah terlibat, berbagi, dan membuat konten, termasuk blog, jejaring sosial, wiki, forum, dan dunia virtual. Twitter adalah salah satu media sosial yang paling populer di kalangan pengguna internet, karena kesederhanaan dan kemudahan penggunaannya dan pengguna dapat dengan bebas mengekspresikan pandangan atau pendapat mereka.

## 2.4 Pengumpulan data

Pengumpulan data dilakukan melalui proses eksplorasi. Crawl merupakan metode penting untuk mengumpulkan data dan memonitor perkembangan internet yang pesat [7]. Penambahan data dari media sosial Twitter menggunakan Google Colabs dan Aggregated Tweet Library akan disimpan dalam file Excel (.xlsx), yang kemudian akan diimpor ke database MySQL. Dataset atau data yang digunakan dalam penelitian ini berasal dari Twitter sebagai data teks tweet yang diperoleh antara tanggal 20 Maret 2023 sampai dengan 29 Mei 2023 dengan total 509 data. Atribut data yang dikumpulkan melalui proses penambahan meliputi: id tweet, teks (tweet), dibuat\_at, nama pengguna, akan disimpan dalam file Excel (.xlsx). Gambar 2 merupakan langkah pengumpulan data.

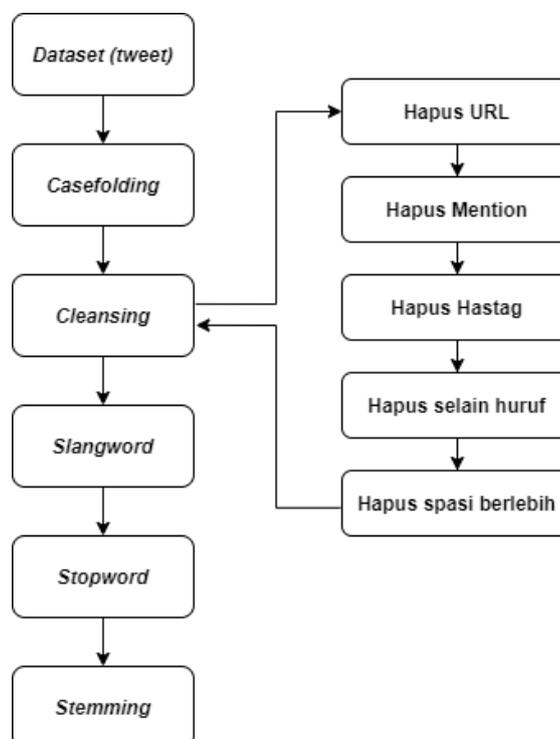


Gambar 2. Alur Pegumpulan data

## 2.5 Preprocessing

*Preprocessing* adalah langkah pertama yang paling penting dalam analisis sentimen ketika menggunakan data Twitter, karena mempengaruhi kinerja classifier [8]. Proses tersebut dilakukan pada tahap *preprocessing*, diantaranya: melipat, membersihkan, mengoreksi kata gaul, menghentikan penghapusan kata dan radikalisasi. Ilustrasi proses langkah *preprocessing* data ditunjukkan pada Gambar 3.

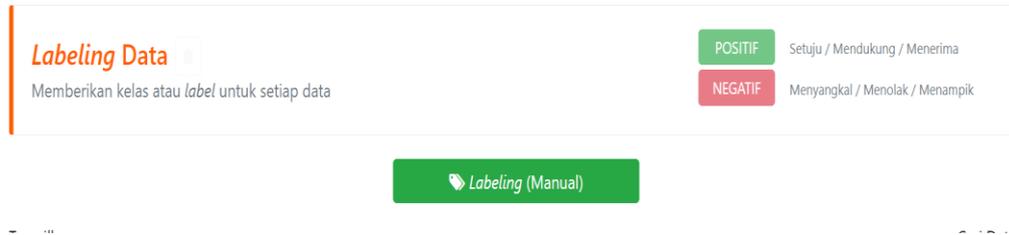
- Cleansing* adalah proses menghilangkan tanda baca, angka, simbol, atau URL dan tautan nama pengguna dalam teks [8].
- Konversi slang adalah proses mengubah kata tidak baku menjadi kata baku, yang dilakukan dengan menggunakan kamus kata slang. Kata-kata dalam kamus slang kemudian diubah menjadi kata baku. Contoh kata slang adalah baper, sotoy, mager dan lain-lain [9].
- Penghapusan kata berhenti menghapus kata-kata yang tidak perlu dari teks. Seperti kata ‘bisa’, ‘tutup’, ‘bisa’ dan ‘tidak bisa’ [8].
- Stemming* akan mengubah kata dalam tweet ke bentuk dasarnya sehingga dapat mengurangi ragam fitur dengan makna yang sama, namun karena imbuhan yang melekat pada kata tersebut menyebabkan perbedaan makna kata tersebut [8].



Gambar 3. Alur *Preprocessing*

## 2.6 Pelabelan

Pelabelan merupakan langkah penandaan dokumen (tweet) yang diperoleh dari hasil koleksi Twitter yang telah melalui proses seleksi data [10]. Pada penelitian ini, proses pelabelan akan memberikan setiap tweet sebuah kelas baik positif maupun negatif (2) kelas (Gambar 4) yang dilakukan pelabelan secara manual dengan (3) orang yang melakukan pelabelan dan (1) pemverifikasi data.



Gambar 4. Labelling Manual

## 2.7 Pembagian data

Pada tahap berbagi data, tweet yang diberi tag akan dibagi menjadi dua (2) bagian diantaranya: data uji dan data latih. Dekomposisi data dilakukan dengan membagi kumpulan data menjadi 90% pelatihan dan 10% pengujian.

## 2.8 Pemodelan

Pemodelan adalah proses menghasilkan pengetahuan dari data pelatihan yang tersedia. Data latih yang digunakan sebagai model dipilih dengan menggunakan teknik quota sampling. Sampling kuota adalah teknik pengambilan sampel yang menentukan jumlah sampel dari suatu populasi dengan karakteristik atau kriteria tertentu hingga kuota yang diinginkan tercapai [11].

## 2.9 Pengujian dan evaluasi

Pengujian model menggunakan data uji [12]. Dalam penelitian ini, pengujian dilakukan pada model berdasarkan data eksperimen yang tersedia. Hasil tes ini akan dievaluasi menggunakan confounding matrix untuk mengukur akurasi, presisi, dan recall.

K-Nearest Neighbor (KNN) adalah salah satu algoritma yang digunakan dalam masalah klasifikasi. Prinsip operasi K-NN adalah mencari jarak terpendek antara data yang akan dievaluasi dengan tetangga terdekatnya. Dalam pencarian data, algoritma K-Nearest Neighbor (KNN) merupakan salah satu algoritma yang paling mudah untuk dipecahkan. sering menghasilkan hasil yang kompetitif dan signifikan. Untuk menghitung jarak digunakan Euclidean [13]. Rumus jarak Euclidean sesuai persamaan (1) berikut ini:

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

Pada penelitian ini, akurasi, presisi, dan daya ingat diuji ketika mengimplementasikan algoritma K-Nearest Neighbor (KNN) pada prediksi label untuk data eksperimen. Selain akurasi, presisi, dan recall, tes pada penelitian ini juga dapat menguji nilai K untuk varian yang telah ditentukan yaitu K = 3, K = 5, K = 7, K = 9, dan K = 11. Selanjutnya setelah menghitung jarak setiap data pembelajaran, dicari data pembelajaran dengan nilai jarak terkecil (tetangga terdekat) yang sama dengan nilai K yang telah ditentukan. Proses terakhir setelah menemukan data tetangga terdekat adalah query. Voting adalah untuk menentukan kelas atau label dari data uji.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Dataset

Pengumpulan set data dilakukan mulai 20 Maret 2023 hingga 29 Mei 2023. Sebanyak 509 tweet tercatat. Properti kepemilikan meliputi: tweet\_id, teks lengkap (tweet), dibuat di, nama pengguna.

### 3.2 Pemrosesan awal data

Seperti yang telah dijelaskan pada bagian sebelumnya, data asli yang diperoleh dari jejaring sosial Twitter telah diproses sebelumnya. Langkah-langkah *preprocessing* yang dilakukan dalam penelitian ini adalah *case folding*, *cleaning*, *slang change*, *stop word removal*, dan radikalisasi. Tabel 1 menunjukkan prosedur langkah *preprocessing* data.

**Tabel 1.** Data Pengolah Kata

Tahap	Input	Output
<i>Case Folding</i>	SERATUS RIBUKU MELAYANG kagak guna ni ada tilang elektronik pulici kaya kagak ada kerjaaaan aje bikin tilangan pas mau lebaran kesal aku ni huft	seratus ribuku melayang kagak guna ni ada tilang elektronik pulici kaya kagak ada kerjaaaan aje bikin tilangan pas mau lebaran kesal aku ni huft
<i>Cleansing</i>	Oh! Suami ketahuan menipu penumpang wanita lain untuk mendapatkan e-tiket <a href="https://t.co/xz5sfxFPn">https://t.co/xz5sfxFPn</a>	Oh! Suami ketahuan menipu penumpang wanita lain untuk mendapatkan e-tiket
<i>Merubah slang word</i>	Global People! lo lebih setuju ada tilang manual atau elektronik aja sih? A.ADA TILANG MANUAL @ajiSABHA	Global People! kamu lebih setuju ada tilang manual atau elektronik aja sih? A.ADA TILANG MANUAL @ajiSABHA
<i>Menghapus stop word</i>	Tilang manual atau elektronik sama saja. Tilang manual sebaiknya hanya focus pada pelanggaran kasat mata saja terutama yang tidak dilengkapi plat nomor, lampu dan knalpot tidak standart, tidak pakai helm yang standart dan yang melanggar rambu	Tilang manual atau elektronik sama saja. Tilang manual sebaiknya hanya focus pada pelanggaran kasat mata saja terutama yang tidak dilengkapi plat nomor, lampu dan knalpot tidak standart, tidak pakai helm standart dan melanggar rambu
<i>Stemming</i>	Ditlantas Polda Jabar Akan Memberlakukan Tilang Manual	Ditlantas Polda Jabar Akan laku Tilang Manual

### 3.3 Pengujian dan Evaluasi

Selama fase pengujian dan evaluasi penelitian ini, model pelatihan dan data eksperimen digunakan. Berdasarkan model pelatihan yang terdiri dari 459 data pelatihan dengan karakteristik sebagai berikut: 81 tweet positif dan 81 tweet pelatihan negatif dan 50 data uji. Tabel 2 berikut merupakan contoh hasil prediksi algoritma KNN dengan nilai K=5.

**Tabel 2.** Data Sampel Prediksi

No	Tweet	Label Aktual	Label Prediksi
1	@TMCPoldaMetro Nah begitu Pak diatur lalu lintasnya biar jalan gak macet dan gak ada yg melawan arah,itu baru namanya kerja melayani masyarakat. Kalau cuma duduk didalam mobil mendata tilang elektronik pengendara yg melawan arah itu sih kalau menurut saya bukan kerja tapi ngadem	Positif	Positif
2	Bagaimana fakta lebih detail dan apa saja penindakan yang dilakukan oleh tim khusus? Yuk Kriuuukers, simak infografis berikut! \n\n#TilangManual #Penilangan #TilangElektronik #Polisi #Indonesia #TimKhusuPenilangan #Razia #Berkendara	Positif	Positif
3	@detikcom Tilang manual atau elektronik sama saja. Tilang manual sebaiknya hanya fokus pada pelanggaran kasat mata saja terutama yang tidak dilengkapai plat nomor, lampu dan knalpot tidak standart, tidak pakai helm yang standart dan yang melanggar rambu.	Positif	Negatif
4	Keduanya punya kelemahan dan kelebihan masing-masing. \n\nTapi, kalau boleh milih lebih baik tilang manual atau elektronik nih Roker Mania? \n\n#RokerMania #kereta #krl #commuterline #krlmania #angker #anakkereta #roker #rombongankereta #infokrl #tilangelektronik #tilangmanual <a href="https://t.co/LjJRcoY7">https://t.co/LjJRcoY7</a>	Positif	Negatif
5	@asumsico Kalo misal pengendara nya salah, tapi polisi nya gak berlisensi gimana? Bisa ttep tilang atau cmn foto plat doang trus nanti lanjut tilang elektronik?	Positif	Negatif
n	.....	.....	.....
50	Pemberlakuan tilang elektronik (Electronic-Traffic Law Enforcement/E-TLE) ternyata tidak lantas membuat sebagian pengguna jalan tertib berlalu lintas. Tilang manual terpaksa digelar kembali untuk menindak para pelanggar #Metro #AdadiKompas\n\n <a href="https://t.co/bDij03PUCo">https://t.co/bDij03PUCo</a>	Negatif	Negatif

Pada Tabel 2 data prediksi sampel, kolom label sebenarnya sesuai dengan data label yang diperoleh melalui proses pelabelan, sedangkan label prediksi sesuai dengan data label yang dihasilkan oleh proses pelabelan. Hasil prediksi keseluruhan dari data eksperimen (50 data tweet) kemudian diplot dalam matriks konfusi. Representasi dari matriks konfusi untuk K=5 yang terbentuk pada Tabel 3 dapat dilihat di bawah ini.

**Tabel 3.** Nilai Uji Confusion Matrix K = 5

		Nilai aktual	
		Positif	Negatif
Nilai prediksi	Positif	33	6
	Negatif	10	1

Berdasarkan Tabel 3, uji matriks konfusi K=5 menunjukkan nilai akurasi, presisi, dan recall yang diperoleh pada Tabel 4 di bawah ini.

**Tabel 4.** Nilai Uji K=5

Pengujian		
Akurasi	= 34/50	0.68 (68%)
Presisi	= 33/39	0.85 (85%)
Recall	= 33/43	0.77 (77%)

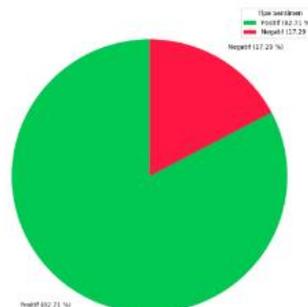
Pengujian di atas dilakukan berulang kali dengan variasi nilai K yang berbeda-beda, sehingga hasil pengujian secara keseluruhan dapat dilihat seperti pada Tabel 5 berikut ini.

**Tabel 5.** Hasil Tes dan Evaluasi

	K=3	K=5	K=7	K=9	K=11
Akurasi	0.62	0.68	0.48	0.6	0.6
Presisi	0.83	0.85	0.79	0.85	0.85
Recall	0.7	0.77	0.53	0.65	0.65

Berdasarkan Tabel 5 di atas dapat dilihat hasil percobaan. Seperti yang dapat dilihat, hasil pengujian menunjukkan bahwa algoritma KNN dapat mencapai nilai eksperimen tertinggi saat menggunakan K=5, dengan akurasi 68%, akurasi 85%, dan akurasi recall 77%. Sementara itu, hasil analisis psikologis dari 509 kicauan menunjukkan arah (persepsi) masyarakat Indonesia memiliki tren positif sebesar 82,71%, sedangkan sentimen negatif sebesar 17,29% pada periode Maret hingga Mei 2023 (Gambar 5).

*Pie Chart* persentase sentimen pada 509 Data tweet.  
421 tweet positif (82.71 %) & 88 tweet negatif (17.29 %)



**Gambar 5.** Grafik Lingkaran Persentase Sentimen

#### 4. KESIMPULAN

Berdasarkan kumpulan data yang telah diolah, outlook masyarakat Indonesia (psikologis) terhadap e-tiket cenderung sentimen positif sebesar 82,71%, sedangkan sentimen negatif sebesar 17,29% untuk periode Maret hingga Mei 2023. Langkah-langkah utama dari penelitian ini meliputi: eksplorasi, *preprocessing*, pelabelan, pemodelan, pembagian data, dan klasifikasi K-Nearest Neighbor (KNN). Langkah *preprocessing* yang baik akan

menjadi faktor penentu dalam membentuk hasil yang optimal untuk langkah selanjutnya. Menggunakan ekstraksi CountVectorizer dan algoritma K-Nearest Neighbor (KNN) dalam analisis sentimen dapat bekerja dengan baik, dengan skor tes dan peringkat tertinggi dicapai oleh: Akurasi 68%, Akurasi 85%, 77 penarikan kembali dengan K=5. Melakukan proses pelabelan dengan cara manual dengan bantuan ahli atau pakar dalam bidang bahasa. Menggunakan Pustaka atau plugin pemrograman yang dapat meringkas waktu pemrosesan data. Melakukan pembagian data dengan rasio pembagian yang lebih beragam untuk mendapatkan data yang optimal.

## DAFTAR PUSTAKA

- [1] M. Afdal and L. Rahma Elita, "Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, pp. 78–87, 2022.
- [2] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [3] M. S. Alrajak, I. Ernawati, and I. Nurlaili, "Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN)," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 1, no. 2, pp. 110–122, 2020.
- [4] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [5] A. V. Sudiantoro *et al.*, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [6] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [7] N. A. Rakhmawati, R. B. Waskitho, D. A. Rahman, and M. F. A. U. Nuha, "Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation," *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 78–83, 2021, doi: 10.26740/jieet.v5n2.p78-83.
- [8] 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)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [9] N. Nofiyani and W. Wulandari, "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1621, 2022, doi: 10.30865/mib.v6i3.4332.
- [10] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 239, 2020, doi: 10.26418/justin.v8i3.39118.
- [11] M. Priandi and Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 311–319, 2021.
- [12] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [13] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.