

## ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP INFRASTRUKTUR JALAN DI LAMPUNG MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Muhammad Ravi Arjunanto<sup>1\*</sup>, Sejati Waluyo<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 Selatan, Indonesia

Email: <sup>1\*</sup>1911501367@student.budiluhur.ac.id, <sup>2</sup>sejati.waluyo@budiluhur.ac.id

(\* : corresponding author)

**Abstrak**-Terutama di negara berkembang, infrastruktur jalan merupakan faktor penting dalam mendorong pertumbuhan. Selain itu, pemerintah berperan penting dalam pembangunan infrastruktur dengan menyediakan pendanaan. Seperti diketahui, jalan merupakan salah satu bentuk infrastruktur publik; oleh karena itu, keterlibatan dan intervensi pemerintah diperlukan untuk mencapai hasil terbaik dan mendorong pembangunan ekonomi. Efek jangka panjang atau pasca konstruksi jalan juga telah menarik perhatian sejumlah penelitian sebelumnya. Namun di sebagian daerah masih ditemukan jalan yang rusak, contoh nya ada di daerah lampung. Pemerintah daerah lampung diduga lalai dalam pembangunan infrastuktur jalan di daerah lampung, dan sudah bertahun-tahun lama nya dibiarkan rusak seperti itu. Lalu ada seorang pemuda asal lampung yaitu Bima yang telah memViralkan jalan di lampung, dengan tujuan agar pemerintah dapat memperhatikan lagi infrastruktur-infratsuktur di daerah di wilayah indonesia. Setelah viral, pak presiden jokowi dengan para menteri beserta jajarannya meninjau ke lampung untuk melihat secara langsung kondisi jalan rusak di lampung. Diharapkan dengan peninjauan langsung oleh pak presiden jokowi akan berdampak baik bagi masyarakat daerah lampung yang terdampak oleh jalan rusak. Tujuan penelitian ini adalah untuk mengetahui penilaian masyarakat terhadap jalan rusak di Lampung. Metode yang digunakan mengekstrak algoritma klasifikasi CountVectorizer dan K-Nearest Neighbor. Dataset terdiri dari query (tweet) yang mengandung kata kunci "Infrastruktur Lampung" di jejaring sosial Twitter. Analisis terhadap 350 tweet mengungkapkan sentimen positif sebesar 30,29% dan sentimen negatif sebesar 69,71% antara tanggal 7 April hingga 19 Juni 2023. Hasil pengujian terbaik diperoleh dengan menggunakan K=5, dengan akurasi sebesar 60%, presisi sebesar 36 %, dan penarikan 50%.

**Kata Kunci:** twitter, infrastruktur lampung, *countvectorizer*, *k-nearest neighbor*

## SENTIMENT ANALYSIS OF TWITTER SOCIAL MEDIA ON ROAD INFRASTRUCTURE IN LAMPUNG USING *K-NEAREST NEIGHBOR* ALGORITHM

**Abstract**-Especially in developing countries, road infrastructure is an important factor in driving growth. In addition, the government plays an important role in infrastructure development by providing funding. As is known, roads are a form of public infrastructure; Therefore, government involvement and intervention is necessary to achieve the best results and promote economic development. The long-term or post-road construction effects have also attracted the attention of a number of previous studies. But in some areas there are still damaged roads, an example is in the Lampung area. The Lampung local government is suspected of negligence in the construction of road infrastructure in the Lampung area, and has been left damaged for many years. Then there is a young man from Lampung, Bima who has viralized roads in Lampung, with the aim that the government can pay attention again to infrastructure in areas in Indonesian territory. After going viral, President Jokowi with his ministers and ranks visited Lampung to see firsthand the condition of damaged roads in Lampung. It is hoped that with a direct review by Mr. President Jokowi, it will have a good impact on the people of the Lampung area affected by damaged roads. The purpose of this study was to determine the community's assessment of damaged roads in Lampung. The method used extracts the classification algorithms CountVectorizer and K-Nearest Neighbor. The dataset consists of queries (tweets) containing the keyword "Lampung Infrastructure" on the social network Twitter. Analysis of 350 tweets revealed positive sentiment of 30.29% and negative sentiment of 69.71% between April 7 and June 19, 2023. The best test results were obtained using K=5, with an accuracy of 60%, precision of 36%, and a draw of 50%.

**Keywords:** twitter, lampung infrastructure, *countvectorizer*, *k-nearest neighbor*

## 1. PENDAHULUAN

Keberadaan infrastruktur jalan sangat berperan penting dalam memperlancar pembangunan, khususnya di negara-negara yang sedang dalam proses pertumbuhan dan kemajuan ekonomi. Pemerintah berperan penting dalam memfasilitasi investasi untuk pembangunan infrastruktur. Diakui secara luas bahwa jalan raya merupakan salah satu bentuk infrastruktur publik, sehingga memerlukan keterlibatan dan tindakan pemerintah untuk memberikan hasil optimal yang mendorong kemajuan ekonomi. Untuk itu, penelitian ini bertujuan untuk mengidentifikasi penyebab kerusakan jalan di provinsi Lampung dan memberikan rekomendasi solusi yang tepat. Penelitian ini melibatkan survei dan analisis data yang meliputi pengukuran kerusakan jalan, analisis cuaca, dan pengumpulan informasi dari masyarakat dan pihak terkait

Analisis sentimen, juga dikenal sebagai analisis sentimen dalam konteks bahasa Indonesia, mengacu pada metodologi yang digunakan untuk memastikan cara penyampaian emosi melalui teks tertulis, dan kemudian mengklasifikasikan emosi tersebut ke dalam kategori positif atau negatif [1]. Sudut pandang serupa diartikulasikan dalam referensi [2], di mana analisis sentimen digunakan untuk memahami komentar pengguna di internet dan menjelaskan persepsi mereka terhadap suatu produk atau merek. Sesuai dengan temuan [3], analisis sentimen adalah prosedur komputasi yang digunakan untuk membedakan dan mengkategorikan pandangan, emosi, dan sikap yang diekspresikan dalam data tekstual, biasanya membedakan antara sentimen negatif dan positif. Seperti yang dikatakan sebelumnya, sejumlah besar pengguna internet terlibat dalam praktik mendokumentasikan pengalaman pribadi mereka, mengungkapkan pendapat mereka, dan berbagi berbagai aspek kehidupan mereka. Menyusun daftar lengkap perasaan yang biasa dialami individu, dikategorikan sebagai menyenangkan, netral, atau negatif, dan sering dikomunikasikan dengan cara yang rumit.

Penelitian sebelumnya tentang analisis sentimen mencakup penyelidikan perintis yang menerapkan teknik penambangan teks pada aplikasi Tokopedia, menggunakan algoritma K-Nearest Neighbor. Studi ini melaporkan penggunaan 5.254 contoh data, yang dipartisi menggunakan validasi silang K-fold, dan kemudian dievaluasi keakuratannya. Parameter  $K=200$  sesuai dengan nilai 71,12%. Dalam kajian terpisah, dilakukan analisis terhadap sentimen seputar fenomena “new normal” di Indonesia pada masa Covid-19. Dataset yang digunakan dalam analisis ini terdiri dari 1000 tweet, dengan 811 tweet yang mengekspresikan sentimen positif dan 189 tweet yang mengekspresikan sentimen negatif. Algoritma klasifikasi K-NN, ketika diterapkan dengan nilai  $K=1$ , menunjukkan akurasi sempurna 100% pada set pelatihan. Selain itu, akurasinya mencapai 92,60% saat dievaluasi menggunakan validasi silang 10 kali lipat dan 94,50% saat menggunakan pembagian persentase 80%. Algoritma K-Nearest Neighbor (KNN) digunakan dalam penelitian bernama “Analisis Sentimen Layanan PT. PLN di Jakarta di Twitter”. Penelitian ini melibatkan klasifikasi data menggunakan KNN dengan metrik jarak Euclidean. Studi ini menggunakan nilai  $K = 3$  untuk menentukan jarak antara titik data yang berdekatan. Mempertimbangkan biaya Angka akurat yang tercatat untuk wilayah DKI Jakarta pada Februari 2020 adalah 89,4 sesuai dengan peringkatnya dari total 500 titik data.

Media sosial, berfungsi sebagai platform berbasis internet, memungkinkan individu untuk membuat profil publik atau semi publik di dalam sistem. Melalui platform ini, pengguna dapat terhubung dengan individu lain dalam daftar pilihan, dengan siapa mereka berbagi hubungan. Selain itu, pengguna dapat menjelajahi dan membaca dengan teliti daftar pengguna lain dan hubungan mereka yang terkait. Media sosial sebagai platform bagi pengguna untuk berpartisipasi aktif dalam pembuatan dan pertukaran konten, serta terhubung dengan orang lain melalui komunitas online.

*Text mining* adalah penemuan pengetahuan dalam basis data tekstual, bisa juga disebut penambangan teks atau pencarian data [mendengar] termasuk minat terhadap pengetahuan yang baru dihasilkan, dipahami sebagai bagian dari proses penggalian atau pencarian tekstual yang sebelumnya tidak diketahui data, agar dapat dipahami, memiliki pola atau pengetahuan potensial dan aktual dari kumpulan data tekstual atau ‘*a large body*’ (kumpulan dokumen yang mencatat penggunaan suatu bahasa dalam bentuk tulisan atau lisan secara utuh dan padat) dan tidak terstruktur.

## 2. METODE PENELITIAN

Perencanaan diagram analisis sentimen adalah gambaran yang menunjukkan alur penelitian yang dilakukan. Alur pencarian bisa dilihat pada gambar 1.

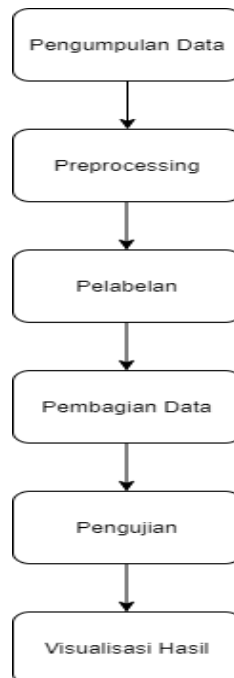
### 2.1 Data Penelitian

Data penelitian yang digunakan pada dalam penelitian ini bersumber dari *twitter* berupa teks (*tweet*) yang diperoleh mulai tanggal 7 April 2023 hingga 19 Juni 2023 sejumlah 350 data. Pemanfaatan teknik crawling sangat penting dalam konteks pengumpulan data dan pemantauan evolusi cepat internet [4]. Data tersebut diperoleh menggunakan pustaka *tweepy* melalui proses *scraping*. Data penelitian yang diperoleh dikumpulkan berdasarkan

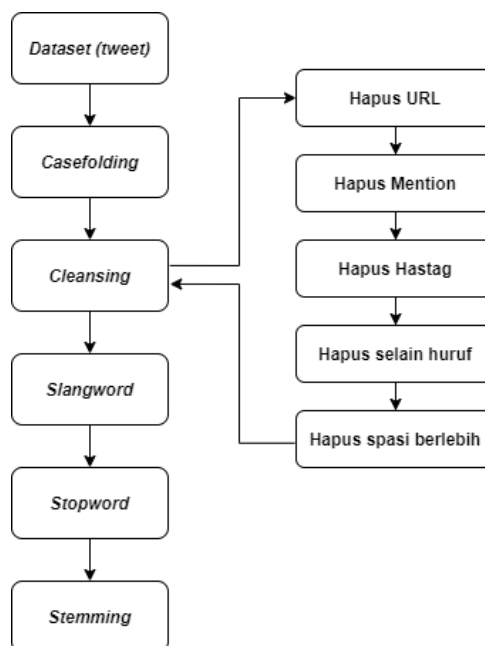
satu parameter kata kunci yang terkait dengan lampung. Data yang dipulihkan kemudian akan disimpan dalam bentuk format CSV. Data akan diberi label secara manual sesuai dengan kelas sentimen positif dan negatif. Pelabelan manual ini digunakan untuk memeriksa keakuratan prediksi yang akan dibuat dengan algoritma K-Nearest Neighbor.

## 2.2 Preprocessing

Tahap awal analisis sentimen saat menggunakan data Twitter adalah preprocessing, yang memiliki kepentingan signifikan karena berdampak langsung pada hasil kinerja klasifikasi [5]. Konversi data teks tidak terstruktur menjadi data teks terstruktur memerlukan prapemrosesan data. Oleh karena itu, penelitian ini akan mencakup beberapa fase preprocessing teks seperti Gambar 2.



Gambar 1. Tahapan Metode

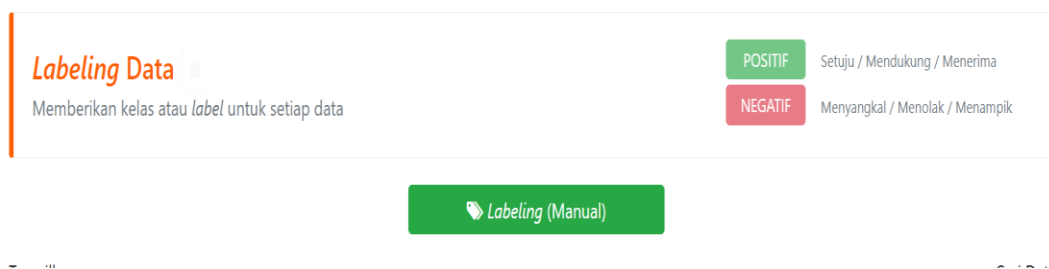


Gambar 2. Tahapan Preprocessing

- “*Preprocessing* merupakan tahapan awal yang paling penting dalam analisis sentimen saat menggunakan data *twitter*, karena berpengaruh terhadap hasil performansi klasifikasi [5].
- Cleaning* merupakan proses untuk menghilangkan tanda baca, angka, simbol, atau *link* URL, dan username dalam teks [5].
- Mengubah *slang word* merupakan proses perubahan 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, messel mager dan lain-lain [6].
- Menghapus *stop word* akan menghilangkan kata-kata yang dianggap tidak penting didalam teks. Seperti kata ‘bisa’, ‘tegas’, ‘mampu’, dan ‘tidak’ [5].
- Stemming *Stemming* akan mengubah kata yang ada pada tweet menjadi bentuk dasarnya sehingga dapat mengurangi variasi fitur-fitur yang memiliki makna yang sama namun karena terdapat imbuhan pada kata tersebut menyebabkan adanya perbedaan makna kata [5]”

### 2.3 Pelabelan

Pelabelan merupakan adalah tahap pemberian label pada dokumen (*tweet*) yang diperoleh dari hasil *twitter* crawling dan telah melewati proses data selection [7]. Pada penelitian ini, proses pelabelan mengklasifikasikan setiap tweet menjadi positif dan negatif. Kelas yang dilakukan dengan melakukan pelabelan manual dengan (2) orang yang melakukan pelabelan dan (1) orang yang melakukan validasi data tersebut.



Gambar 3. Labeling Manual

### 2.4 Pembagian data

Berbagi data mengacu pada praktik mempartisi kumpulan data yang diberikan menjadi dua kelompok berbeda, biasanya disebut sebagai kumpulan pengujian dan pelatihan. Data pelatihan mengacu pada kumpulan data spesifik yang digunakan untuk tujuan membangun dan menginstruksikan model kategorisasi. Pengujian mengacu pada dataset yang akan dievaluasi oleh pengklasifikasi, yang telah dilatih menggunakan data pelatihan tertentu. Prosesnya melibatkan pembuatan model klasifikasi data menggunakan kumpulan data pelatihan yang ditunjuk, diikuti dengan evaluasi kinerja pengklasifikasi menggunakan kumpulan data uji yang terpisah. Biasanya, proporsi data pelatihan terhadap data uji adalah 90:10, artinya 90% data dialokasikan untuk tujuan pelatihan dan 10% sisanya dicadangkan untuk mengevaluasi kinerja model. Pencapaian hasil optimal dalam klasifikasi bergantung pada komposisi data pelatihan. Khususnya, jika data pelatihan mencakup sebagian besar data yang diperlukan untuk pengujian selanjutnya, hasil yang dihasilkan akan dimaksimalkan.

### 2.5 Pemodelan

Pemodelan mengacu pada prosedur sistematis untuk memperoleh informasi atau wawasan dari data pelatih yang ada. Data latih model dipilih melalui pemanfaatan teknik kuota sampling. Sampling kuota adalah strategi pengambilan sampel yang digunakan untuk memastikan jumlah sampel dari suatu populasi yang memiliki sifat atau kriteria tertentu, sampai kuota yang diinginkan tercapai [8].

### 2.6 K-Nearest Neighbor

Teknik K-Nearest Neighbor (KNN) umumnya digunakan dalam tugas klasifikasi. Algoritma K-Nearest Neighbor (KNN) adalah pendekatan langsung yang digunakan untuk mengatasi masalah klasifikasi. Premis operasinya melibatkan penentuan jarak terkecil antara data yang akan dievaluasi dan tetangga terdekat dalam data penelitian. Meskipun sederhana, teknik KNN sering menghasilkan hasil yang kompetitif dan signifikan secara statistik. Untuk menghitung jarak melalui metode Euclidean [9]. Rumus jarak Euclidean sesuai Persamaan (1) adalah sebagai berikut:

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

Keterangan:

- $D(x,y)$  = Jarak antara data uji dengan latih  
 $n$  = Jumlah fitur  
 $X_i$  = Fitur ke- $i$  dalam data uji  
 $Y_i$  = Fitur ke- $i$  dalam data latih

Proses selanjutnya setelah menghitung jarak untuk setiap data latih adalah mencari data latih dengan nilai jarak terkecil (ketetanggaan terdekat) sebanyak nilai  $K$  yang telah ditentukan. Proses akhir setelah ditemukannya data tetangga terdekat adalah pemungutan suara (voting). Voting bertujuan untuk menentukan kelas atau label dari suatu data uji.

## 2.7 Pengujian

Pengujian model menggunakan data uji [10]. Pekerjaan ini melibatkan melakukan tes pada model yang dikembangkan menggunakan data eksperimen yang ada. Evaluasi temuan pengujian akan dilakukan dengan menggunakan matriks perancu seperti Tabel 1 untuk menilai tingkat akurasi, presisi, dan daya ingat.

**Tabel 1.** *Confusion Matrix*

		Nilai Aktual	
		TRUE ( <i>positive</i> )	FALSE ( <i>negative</i> )
Nilai Prediksi	TRUE ( <i>positive</i> )	TP (True Positive)	FP (False Positive)
	FALSE ( <i>negative</i> )	FN (False Negative)	TN (True Negative)

Eksperimen dilakukan untuk menilai kemanjuran dan kinerja model pelatihan dengan menggunakan algoritma yang disarankan, dengan fokus pada evaluasi nilai, akurasi, presisi, dan daya ingatnya. Studi ini melibatkan pelaksanaan pengujian dengan membandingkan data prediksi (khususnya, data langkah klasifikasi) dengan sekumpulan data aktual yang ditunjuk (disebut sebagai pelabelan data hasil). Algoritma K-Nearest Neighbor (KNN) akan digunakan dengan beberapa nilai  $K$ , yaitu  $K=3$ ,  $K=5$ ,  $K=7$ ,  $K=9$ , dan  $K=11$ .

## 3. HASIL DAN PEMBAHASAN

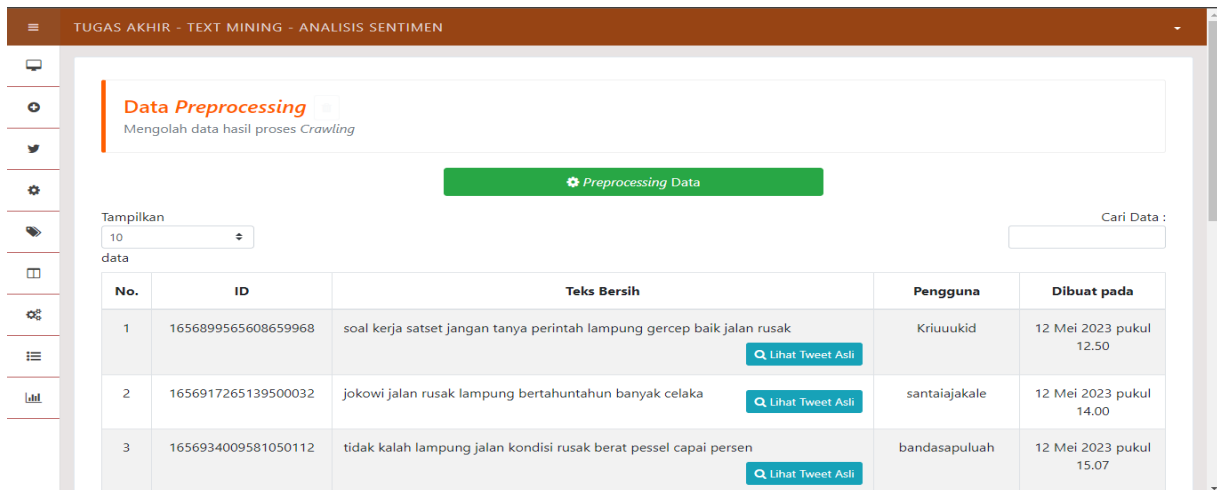
Hasil dari eksperimen ini adalah nilai akurasi algoritma pada kelas prediksi pada data uji. Hasil akurasi ini diperoleh setelah melakukan proses penarikan data dari twitter atau *scraping*, yang di dapat adalah id pengguna, *text* atau komentar, *username* pengguna, *created\_at* atau waktu *upload* komentar, seperti yang ditunjukkan pada Gambar 4. Mulailah dengan mengumpulkan data di Twitter dan memberi label sentimen secara manual. Data yang baru saja diambil dari Twitter masih mengandung banyak ikon dan tautan. Oleh karena itu, perlu masuk ke tahap preprocessing yang ditunjukkan pada Gambar 5 untuk menghilangkan *icon*, *link* dan membuat setiap tweet lebih konsisten.

Ketika semua tweet diproses sebelumnya, itu menghasilkan data bersih yang dapat dimasukkan ke dalam proses selanjutnya, adalah berbagi data yang ditunjukkan pada Gambar 6. Data tersebut dibagi menjadi 2 yaitu data latih dimana setiap kata akan disimpan dalam kantong kata sebagai data latih. Pengujian data akan menjadi pengujian masing-masing algoritma untuk memprediksi kelas afektif dan menggunakan data pelatihan yang tersimpan sebagai acuan untuk memprediksi kelas tersebut. Menggunakan mode 90:10 dimana 90% dari total data adalah data latih dan 10% adalah data uji.

Data uji ini akan diuji dengan algoritma K-Nearest Neighbor untuk memprediksi kelas afektif menggunakan data latih yang disimpan dengan metode word bag sebagai nilai setiap kata. Hasil prediksi akan dibandingkan secara manual dengan pelabelan asli untuk melihat keakuratannya. Bisa dilihat pada Gambar 7.

A	B
1	id text
2	1,67E+18 saya membayangkan gubernur jatim nelpon gubernur riau, protes soal ekspor pasir laut. atau gubernur papua mempersoalkan tambang di sangihe sama gubernur sulut. atau gubernur jabar protes soal jala
3	1,67E+18 Koreksi, di lampung bukan jalan pedesaan yg rusak, jalan kabupaten jalan propinsi @tinggipasaribu @Lembayung071 @masbutet
4	1,67E+18 komunikasi, mendengar... membuka saluran keluhan warga, itu salah satu kelebihan ganjar dibanding calon2 lain. diakui ato tidak terserah...@miskiy28
5	1,67E+18 Halo Mas @masbutet sampeyan benerasi ala WS Rendra kalau Ganjar sudah kerja jungkir balik siang malam, akh ..masa..?
6	1,67E+18 Buktinya setelah viral, presiden langsung tanggap jalan rusak di lampung. Waktu era SBY, desa2 di daerah saya mendapat dana dr usah utk pembangunan jalan dan jembatan dipelosok2 desa, yg dulu teriso
7	1,67E+18 Kapolsek Iptu Suryono berserta Anggota polsek marga tiga melaksanakan giat Bergotong Royong menimbulk jalan yang rusak dan susah untuk dilalui kendaraan roda empat di Jalan Desa Gedung Wani Timu
8	1,67E+18 Jalan tol di Lampung, dulunya hanya angan2 yang di pamerkan setahun sekali di Pameran Pembangunan PKOR, dan pada malam hari tidak ada pemadaman rutin seperti jaman SBY dulu (PLTU Tarahan rusak
9	1,67E+18 "Peristiwa tabrak lari terjadi di Jalan Lintas Sumatera, Tarahan, Lampung Selatan. Merenggut nyawa pria berusia 27 tahun yang baru pulang kerja. Kondisi jalan rusak ditengarai jadi penyebabnya. https://t.t
10	1,67E+18 "Aduh, Mobil Jokowi Kesulitan Lewati Jalan Rusak di Lampung - https://t.co/9g6sYcGAPW #jokowi #prabowo #PrabowoSubianto https://t.co/ULrM1NY9p"
11	1,67E+18 "Aduh, Mobil Jokowi Kesulitan Lewati Jalan Rusak di Lampung - https://t.co/2hHic1YHXW #Prabowo #Jokowi #PrabowoAja https://t.co/u3n5Y11hco"
12	1,67E+18 "Aduh, Mobil Jokowi Kesulitan Lewati Jalan Rusak di Lampung - https://t.co/ok84PYxygO #prabowo #jokowi #PrabowoAja https://t.co/6nixFCVwpQ"
13	1,67E+18 "Aduh, Mobil Jokowi Kesulitan Lewati Jalan Rusak di Lampung - https://t.co/xGyA7L5jD9 #prabowo #jokowi #Prabowoaja https://t.co/hBpblTaUeZ"
14	1,67E+18 "Aduh, Mobil Jokowi Kesulitan Lewati Jalan Rusak di Lampung - https://t.co/dyu0VdDuWV #Prabowo #Jokowi #PrabowoSubianto https://t.co/ulFqw7OW3R
15	1,67E+18 "Anggota Satuan Lalu Lintas Polres Tanggamus, Ps. Kanit Kamsel Briпка Bayu mewakili Kasat Lantas AKP Amсар, S.Sos, mendampingi Dinas Bina Marga dan Bina Konstruksi (BMBK) Provinsi Lampung meninje
16	1,67E+18 Ooo... gak kebanyakan VVIP bareng WowoWiwiThohirðYzE Kebagian bareng engok jalan rusak doang di Lampung! hahaha.
17	1,67E+18 Kepada yg terhormat bapak @jokowi bukan hanya jalan yg rusak di kampung halaman saya tetapi juga sinyal yg sulit menghambat proses belajar saya sebagai mahasiswa, mohon"
18	1,67E+18 "Tak Hanya Cek Jalan Rusak, Arinal Djunaidi juga Pantau Kebutuhan Pokok di Lampung Timur

Gambar 4. Dataset



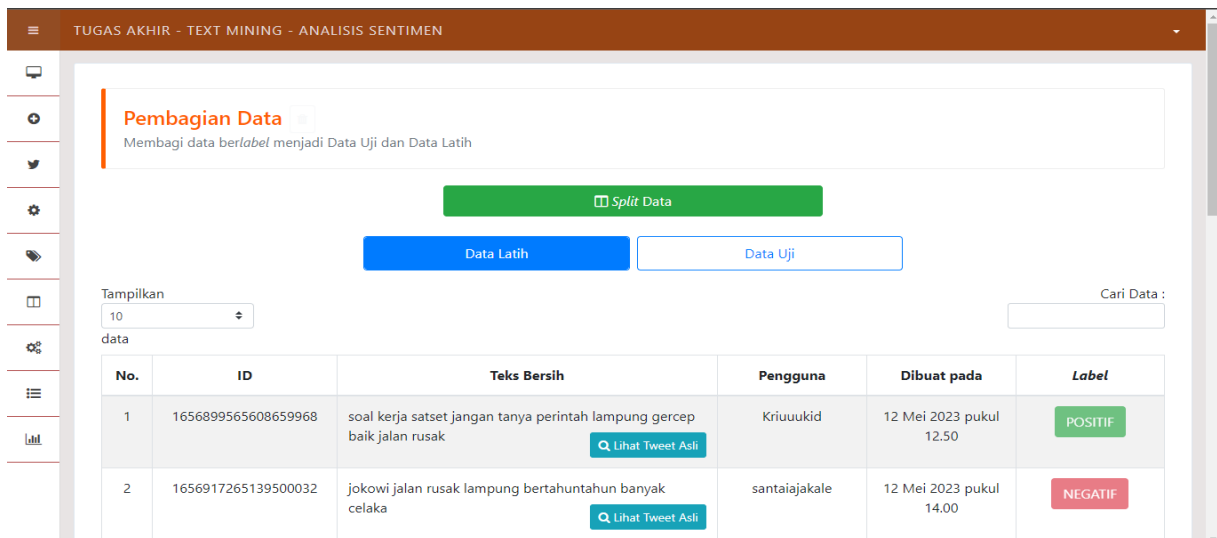
**Data Preprocessing**  
Mengolah data hasil proses Crawling

Preprocessing Data

Tampilkan: 10 data

No.	ID	Teks Bersih	Pengguna	Dibuat pada
1	1656899565608659968	soal kerja satset jangan tanya perintah lampung gercep baik jalan rusak	Kriuuukid	12 Mei 2023 pukul 12.50
2	1656917265139500032	jokowi jalan rusak lampung bertahun-tahun banyak celaka	santaiajakale	12 Mei 2023 pukul 14.00
3	1656934009581050112	tidak kalah lampung jalan kondisi rusak berat pesel capai persen	bandasapuluah	12 Mei 2023 pukul 15.07

Gambar 5. Preprocessing



**Pembagian Data**  
Membagi data berlabel menjadi Data Uji dan Data Latih

Split Data

Data Latih Data Uji

Tampilkan: 10 data

No.	ID	Teks Bersih	Pengguna	Dibuat pada	Label
1	1656899565608659968	soal kerja satset jangan tanya perintah lampung gercep baik jalan rusak	Kriuuukid	12 Mei 2023 pukul 12.50	POSITIF
2	1656917265139500032	jokowi jalan rusak lampung bertahun-tahun banyak celaka	santaiajakale	12 Mei 2023 pukul 14.00	NEGATIF

Gambar 6. Pembagian Data



**Penguujian Data**  
Menguji Model latih menggunakan Data Uji

Tampilkan: 10 data

Cari Data :

No.	Tweet	Sentimen (Aktual)	Sentimen (Prediksi)	Tetangga Terdekat
1	Seorang Mahasiswa bernama Bima mengkritik jalan dan menyindir pihak pemerintah terhadap banyak jalanan yang rusak di kota lampung di sosial media. Kritik tersebut membuat warga internet heboh dan terjadinya digital activism di sosial media	POSITIF	POSITIF (60%)	<a href="#">Detail</a>
2	Di Bengkulu dan Lampung, mengurus jenazah jadi tak mudah gegara jalan rusak dan tak adanya jembatan. Bagaimana ceritanya?https://t.co/mqnpN6gSs1	NEGATIF	NEGATIF (60%)	<a href="#">Detail</a>
3	Mending Abang enak nya speakup tentang hal yang ada di daerah abang aja , kaya Abang Bima dari Lampung terkenal dia skrng hanya karena kritik jalan rusak. Dari pada ngejek ngehina Palsong ga ada untungnya buat Abang	POSITIF	NEGATIF (60%)	<a href="#">Detail</a>

**Gambar 7.** Penguujian Data

Selama fase pengujian dan evaluasi ini, model pelatihan dan data uji digunakan. Berdasarkan model pelatihan yang terdiri dari 315 data pelatihan dengan karakteristik sebagai berikut 96 tweet positif dan 96 tweet pelatihan negatif dan 35 data uji, berikut pada Tabel 2 adalah sampel hasil prediksi algoritma KNN dengan nilai K=7.

**Tabel 2.** Nilai Confusion Matrik Pengujian K=7

		Nilai aktual	
		Positif	Negatif
Nilai Prediksi	Positif	3	8
	Negatif	7	17

Berdasarkan Tabel 2 Uji Confusion Matriks K=7, nilai akurasi, presisi dan *recall* dapat diperoleh pada Tabel 3.

**Tabel 3.** Nilai Pengujian K=7

Pengujian	
Akurasi	= 20/35 0.57 (57%)
Presisi	= 3/11 0.27 (27%)
<i>Recall</i>	= 3/10 0.3 (30%)

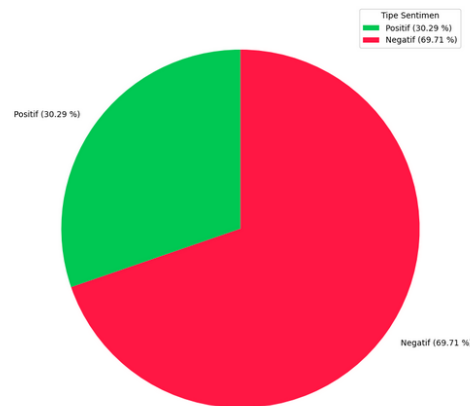
Pengujian di atas dilakukan dengan variasi nilai K yang berbeda, sehingga hasil pengujian secara keseluruhan dapat dilihat seperti pada Tabel 4.

**Tabel 4.** Hasil Pengujian dan Evaluasi

	K=3	K=5	K=7	K=9	K=11
Akurasi	0.53	0.6	0.57	0.63	0.69
Presisi	0.31	0.36	0.27	0.29	0.4
<i>Recall</i>	0.5	0.5	0.3	0.2	0.2

Berdasarkan tabel 4, diketahui hasil pengujian, dapat diketahui bahwa hasil pengujian menunjukkan bahwa algoritma KNN mampu memperoleh nilai pengujian tertinggi menggunakan nilai K=7, dengan akurasi 57%, presisi 27%, dan *recall* 30%. Sementara itu hasil analisis sentiment terhadap 350 tweet menunjukkan arah pandangan (sentiment) masyarakat Indonesia cenderung negatif 69.71%, sementara sentimen positif sebesar 30.29 % pada periode april hingga juni 2023.

*Pie Chart* persentase sentimen pada 350 Data *tweet*.  
106 *tweet* positif (30.29 %) & 244 *tweet* negatif (69.71 %)



Gambar 8. *Pie Chart*

#### 4. KESIMPULAN

Berdasarkan hasil dan evaluasi dari dataset dan algoritma yang digunakan dalam penelitian ini, yang mencakup sampel 350 *tweet*, dapat disimpulkan bahwa sentimen yang diungkapkan oleh masyarakat Indonesia terhadap infrastruktur jalan yang memburuk dapat dilihat. Tingkat kerusakan di Lampung menunjukkan kecenderungan yang dominan negatif, dengan penurunan sebesar 69,71%. Selama periode April hingga Juni 2023, proporsi sentimen positif tercatat sebesar 30,29%. Pekerjaan ini mencakup beberapa tonggak penting, yaitu *scraping*, *preprocessing*, *labeling*, *modeling*, *data sharing*, dan *klasifikasi K-Nearest Neighbor (KNN)*. Kualitas langkah *preprocessing* memainkan peran penting dalam mempengaruhi pencapaian hasil yang optimal pada langkah selanjutnya.

#### DAFTAR PUSTAKA

- [1] M. Afdal and L. R. Elita, "Penerapan *Text Mining* Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, p. 78, 2022, doi: 10.24014/rmsi.v8i1.16595.
- [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] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.*, vol. 2, no. 2, pp. 311–319, 2021.
- [9] 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.
- [10] 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.