

Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor

by Fathoni Fathoni

Submission date: 05-May-2023 01:55PM (UTC+0700)

Submission ID: 2084868862

File name: belum4.pdf (460.56K)

Word count: 5067

Character count: 30945



Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode *Improved K-Nearest Neighbor*

Siti Nur Arafah*, Fathoni

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹*09031281823145@student.unsri.ac.id, ²fathoni@unsri.ac.id

Email Penulis Korespondensi: 09031281823145@student.unsri.ac.id

Abstrak—LRT merupakan sistem transportasi cepat yang berkelanjutan, yang dibangun untuk mengatasi masalah kemacetan di kota Palembang. Agar dapat menarik minat masyarakat untuk beralih menggunakan transportasi umum dibandingkan dengan transportasi pribadi salah satunya dengan meningkatkan kualitas pelayanan yang diberikan. Analisis sentimen digunakan untuk mengelompokkan opini positif dan negatif pada pengguna jasa transportasi LRT Kota Palembang. Selain melakukan penarikan data melalui crawling data dan data tweet, peneliti juga melakukan penyebaran kuesioner. Dalam melakukan proses klasifikasi analisis sentimen, penelitian ini menggunakan metode *Improved K-Nearest Neighbor* yang merupakan modifikasi dari metode *K-Nearest Neighbor*. Hasil penelitian ini merupakan data latih dan uji terhadap 1617 record data yang menghasilkan akurasi tertinggi sebesar 74.07% pada 90% data latih dan 10% data uji, dengan nilai presisi 70%, recall 56% dan f-1 score 59%, sedangkan akurasi terendah dengan nilai akurasi 63.04% pada 50% data latih dan 50% data uji, dengan presisi 44%, recall 42% dan f-1 score 42%.

Kata Kunci: Analisis Sentimen; K-Nearest Neighbor; Light Rail Transit (LRT)

Abstract—The LRT is a sustainable fast transportation system, which was built to overcome the congestion problem in the city of Palembang. In order to attract people's interest to switch to using public transportation compared to private transportation, one of them is by improving the quality of services provided. Sentiment analysis is used to classify positive and negative opinions on users of Palembang City LRT transportation services. In addition to retrieving data through crawling data on tweet data, the researchers also distributed questionnaires. In conducting the classification process of sentiment analysis, this study uses the *Improved K-Nearest Neighbor* method which is a modification of the *K-Nearest Neighbor* method. The results of this research are testing and training data on 1617 data records and the highest accuracy of 74.07% on 90% training data and 10% testing data, with 70% precision, 56% recall and 59% f-1 score, while the lowest accuracy with an accuracy of 63.04% on 50% training data and 50% testing data, with 44% precision, 42% recall and 42% f-1 score

Keywords: Sentiment Analysis; K-Nearest Neighbor; Light Rail Transit (LRT)

1. PENDAHULUAN

Sistem transportasi di Indonesia saat ini telah berkembang pesat dan memberikan mobilitas yang tinggi kepada masyarakat. Berdasarkan Sensus penduduk tahun 2020 Sumatera Selatan, Kepadatan Penduduk di Kota Palembang mencapai hingga 1.7 juta jiwa, yang dimana mengakibatkan kemacetan lalu lintas di Kota Palembang semakin tahunnya semakin parah sebesar 6 kali lipat. Transportasi berfungsi sebagai sarana penghubung antar tempat maupun wilayah, untuk membuat masyarakat beralih menggunakan transportasi umum masyarakat akan menuntut pelayanan dan fasilitas yang lebih baik sesuai dengan kebutuhannya. Sebagai salah satu contoh transportasi umum di Kota Palembang yaitu *Light Rail Transit* (LRT). LRT merupakan sistem transportasi cepat berkelanjutan yang dibangun untuk mengatasi masalah kemacetan di kota Palembang[1]. LRT dioperasikan oleh PT. Kereta Api Indonesia Divisi Regional III Palembang, yang memiliki beberapa jalur layanan, yang dimulai dari stasiun Bandara Internasional Sultan Mahmud Badaruddin II sampai Jakabaring *Sport City*[2]. Dilihat dari tingkat kemacetan Kota Palembang, masih banyak masyarakat yang menggunakan kendaraan pribadi dan belum beralih menggunakan LRT sebagai alat transportasi umum. Hal ini perlu dipertimbangkan lebih lanjut, untuk mengetahui pendapat masyarakat mengenai apa yang dirasakan ketika menggunakan LRT Palembang, sehingga informasi yang didapatkan bisa menjadi standar dan solusi dalam pengambilan keputusan pemerintah untuk meningkatkan pelayanan dan kualitas pada LRT kota Palembang.

Twitter merupakan layanan dari jejaring media sosial, yang memungkinkan penggunanya untuk dapat berinteraksi dengan pengguna lainnya dalam sebuah pesan yang berbasis teks hingga 280 karakter (tweets)[3]. LRT merupakan salah satu transportasi yang banyak dibahas di media sosial Twitter. Twitter menyediakan data yang tersedia melalui Twitter API, sehingga memudahkan proses pengumpulan tweets dalam jumlah besar[4]. Pada pengguna twitter API, tweets akan diolah menjadi data untuk pencarian sesuai keywords, data tersebut akan diproses menggunakan analisis sentimen, yang apabila diproses dengan baik dan benar dapat menghasilkan informasi yang dapat berguna dalam pengambilan keputusan dan meningkatkan persepsi seorang akan sesuatu hal yang baru diketahui. Selain melakukan penarikan data melalui crawling data pada data tweet, peneliti juga melakukan penyebaran kuesioner. Alasan peneliti melakukan penyebaran kuesioner, peneliti ingin melihat respon masyarakat secara langsung terhadap penggunaan fasilitas dan layanan LRT Palembang, dengan menggunakan kuesioner peneliti bisa mendapatkan reaksi dan opini masyarakat yang lebih beragam, serta mendapatkan record data yang lebih banyak dari proses crawling data pada twitter.



Analisis sentimen atau *opinion mining* adalah proses otomatis untuk memahami, mengekstraksi, dan memproses atau mengolah data berbe¹⁷k tekstual untuk memperoleh informasi sentimen yang terkandung pada kalimat opini[5]. Pada Penelitian ini, Analisis sentimen digunakan untuk mengelompokkan opini positif, negatif dan netral pada peng²⁵a jasa transportasi LRT Kota Palembang. Dalam melakukan proses klasifikasi analisis sentimen, dilakukan menggunakan metode *Improved K-Nearest Neighbor* yang merupakan modifikasi dari metode *K-Nearest Neighbor*. Modifikasi dilakukan karena ada perubahan penentuan nilai-k, dimana penentuan nilai-k tetap dilakukan hanya saja setiap kategori memiliki nilai-k yang berbeda[6]. Metode ini bersifat fleksibel dan tangguh terhadap data yang *noisy* serta efektif digunakan dengan jumlah data latih yang besar.

Salah satu penelitian yang pernah dilakukan oleh Okfalisa tahun 2017 yaitu dengan melakukan perbandingan antara metode KNN dan metode *Improved KNN* pada data *Conditional Cash Transfer Implementation Unit* (Unit Pelaksana Program Keluarga Harapan) dengan 7395 *record* data. Penelitian tersebut memiliki hasil yaitu dengan menggunakan metode K-NN mendapatkan akurasi tertinggi 94.95% dengan rata-rata akurasi selama pengujian 93.94% dan akurasi tertinggi dengan menggunakan metode *Improved KNN* mendapatkan akurasi tertinggi 99.51% dengan rata-rata akurasi 99.20% [7].

Penelitian analisis sentimen lainnya yaitu oleh Daeli dan Adiwijaya pada tahun 2020, penelitian tersebut menggunakan objek *review* film pada Twitter dengan menggunakan metode *K-Nearest Neighbor*. Hasil dari penelitian tersebut memiliki hasil *accuracy* sebesar 96.08% pada nilai $k=3$ [8].

Penelitian lain mengenai analisis sentimen dilakukan oleh Salam pada tahun 2018, dengan menggunakan objek komentar pada jasa ekspedisi barang J&T *express* Indonesia yang diambil dari sosial media *facebook* dengan menggunakan metode KNN. Penelitian tersebut bertujuan untuk mengetahui persentase sentimen positif dan negatif dari pengguna *facebook* dengan metode *K-Nearest Neighbor*, adapun hasil dari pengujian yang diperoleh yaitu nilai *accuracy* sebesar 79.21% dengan uji coba sebanyak 6 kali [9].

Penelitian Analisis Sentimen yang dilakukan oleh Deviyanto dan Wahyudi tahun 2018 ²² ang Pilkada DKI 2017 pada data twitter, dengan menggunakan jumlah 2000 data dan menerapkan metode algoritma KNN dengan ²² bobot kata TF-IDF dan fungsi *Cosine Similarity*. Penelitian tersebut memperoleh nilai akurasi terbesar 67.2% ketika $k=5$, presisi tertinggi 56.94% ketika $k=5$, dan ¹¹ 78.24% dengan $k=15$ [10].

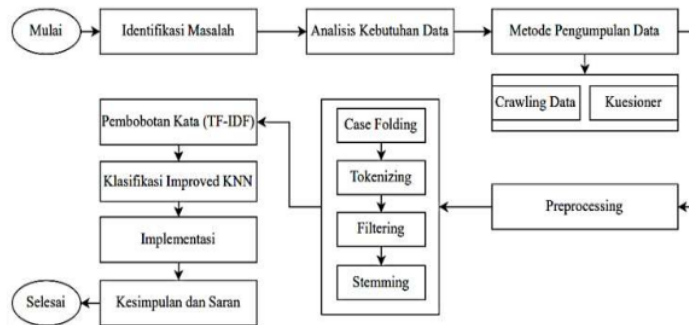
Adapun penelitian tentang analisis sentimen lainnya, yaitu pada ulasan peng⁷na aplikasi bibit dan bareksa dengan algoritma KNN yang dilakukan oleh Putra dan Juanita tahun 2021. Berdasarkan ⁷asil yang diperoleh dari tahapan dengan menggunakan algoritma *k-nearest neighbors* dan perbandingan data 60:40 untuk data *training* dan data *testing*, maka nilai akurasi, *precision* dan *recall* yang dihasilkan dari tiap aplikasi yaitu untuk bibit 85.14% , 91.91%, dan 76.44% sedangkan untuk bareksa yaitu 81.70% , 87.15%, dan 75.73% [11].

Berdasarkan latar belakang dan penelitian terkait, didapatkan rumusan masalah pada penelitian ini, yaitu bagaimana hasil dari klasifikasi sentimen masyarakat terhadap LRT Kota Palembang dan dengan menerapkan metode *Improved K-Nearest Neighbor* (KNN) berapa hasil nilai akurasi yang diperoleh dari sentimen yang didapatkan. Adapun tujuan dari penelitian ini untuk menerapkan metode *Improved KNN* pada bidang analisis sentimen, untuk mengetahui hasil klasifikasi sentimen dari masyarakat terhadap LRT Kota Palembang, serta untuk mengetahui nilai akurasi yang dihasilkan dengan menerapkan metode *Improved KNN* pada analisis sentimen masyarakat terhadap LRT Kota Palembang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah acuan, pedoman dan tahapan yang diterapkan pada sebuah penelitian untuk mencapai tujuan dari penelitian tersebut. Tahapan penelitian memiliki rencana kerja yang sistematis sehingga hasil yang didapatkan dapat sesuai dengan yang diharapkan. Oleh karena itu ³perluan tahapan yang tersusun. Gambar 1. merupakan tahapan yang dilakukan pada penyusunan penelitian *Sentiment Analysis* Pada Masyarakat Terhadap *Light Rail Transit* (LRT) Kota Palembang menggunakan Metode *Improved K-Nearest Neighbor*.



Gambar 1. Tahapan Penelitian

Penjelasan terhadap Gambar 1. Tahapan penelitian *Sentiment Analysis* pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode *Improved K-Nearest Neighbor*.

1. Tahap pertama pada penelitian ini adalah proses mengidentifikasi masalah terhadap penelitian analisis sentimen pada masyarakat terhadap LRT Kota Palembang,
2. Tahap kedua, melakukan analisa kebutuhan data apa saja yang akan digunakan pada penelitian ini, mulai dari data yang akan diolah sampai *tools* atau alat bantu apa saja yang digunakan,
3. Tahap selanjutnya, melakukan proses pencarian referensi terhadap penelitian terdahulu atau studi literatur, dengan cara mencari referensi penelitian sejenis atau penelitian terdahulu,
4. Tahapan keempat adalah pengumpulan data dengan proses *crawling* data dan melakukan survey melalui kuesioner dengan target penyebaran pengguna LRT Palembang,
 - a. Pertama, peneliti mengambil data dengan menarik sampel data dari *twitter* dengan kata kunci "LRT" atau "LRT Palembang" atau "LRT Sumsel" dengan menggunakan media bantuan *tools python 3*.
 - b. Kedua, peneliti menyebarkan kuesioner kepada pengguna LRT Palembang untuk mendapatkan opini terkait penggunaan LRT baik fasilitas maupun pelayanan yang ada pada LRT Palembang.
5. Tahap kelima, melakukan pemrosesan data dengan menggunakan tahapan proses *text mining* terhadap data sentimen yang sudah dikumpulkan, mulai dari proses *preprocessing* yang terdiri dari *case folding*, *tokenizing*, *filtering*, dan *stemming* data pada analisis sentimen masyarakat terhadap LRT Kota Palembang,
6. Tahapan selanjutnya adalah melakukan proses pembobotan kata menggunakan TF-IDF, data yang telah melalui proses *text mining* pada penelitian, akan dilakukan proses pembobotan kata,
7. Tahap selanjutnya adalah mengelola data dengan cara mengklasifikasi data yang sudah diproses pada tahap sebelumnya dengan menggunakan alat bantu *Python 3* serta dengan menggunakan Metode *Improved K-Nearest Neighbor*,
8. Selanjutnya, tahap implementasi olah data yaitu dengan bantuan *tools Python 3* memasukan operator-operator yang sudah disediakan dari analisis sentimen yang sudah didapat,
9. Hasil dari olah data pada *Sentiment Analysis* pada Masyarakat terhadap LRT Kota Palembang menggunakan Metode *Improved K-Nearest Neighbor*, akan didapatkan *output* berupa hasil akurasi dari metode yang digunakan dan klasifikasi sentimen masyarakat terhadap LRT Kota Palembang,
10. Kesimpulan dan Saran berupa hasil akhir dari penelitian *Sentiment Analysis* pada Masyarakat terhadap LRT Kota Palembang menggunakan Metode *Improved K-Nearest Neighbor*.

2.2 Analisis Sentimen

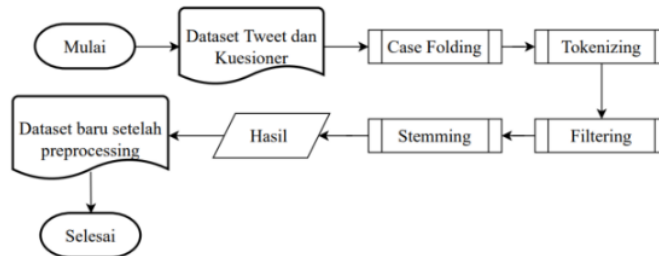
Analisis sentimen dilakukan untuk melihat setiap opini atau kecenderungan opini terhadap suatu isu atau objek dari seseorang, apakah cenderung memiliki pandangan atau opini yang bersifat positif atau negatif[12]. Pada penelitian ini analisis sentimen dibagi menjadi 3 kategori, yaitu sentimen positif, sentimen negatif dan sentimen netral. Sentimen positif adalah kalimat pernyataan yang mengandung dukungan, dan penilaian positif pada sebuah objek, sentimen negatif adalah kalimat pernyataan yang mengandung kalimat kecaman ataupun kritikan pada sebuah objek, sedangkan kalimat netral adalah pernyataan yang tidak mengandung kalimat dukungan maupun kritikan yang membangkitkan emosi pembaca pada sebuah objek[13].

2.3 Preprocessing Text

Text mining merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi, dimana *text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[14]. *Preprocessing Text* merupakan tahapan awal dari *text mining* yang meliputi proses untuk mempersiapkan data tekstual yang akan digunakan agar dapat diproses pada tahapan berikutnya. Proses *Pre-Processing* berfungsi untuk membersihkan data yang terduplikat dan *noise*[15]. Pada tahapan *preprocessing* dilakukan 4 proses lainnya yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*. Adapun alur dari tahapan



preprocessing pada Gambar 2. Setelah data bersih maka dilakukan proses pelabelan secara otomatis dengan *tools python 3* dan penerapan *library NLTK*, *library Sastrawi* dan *library polarity* untuk penentuan nilai *polarity* dari sentimen.



Gambar 2. Alur *Preprocessing Text*

24

2.4 Pembobotan Kata (TF-IDF)

Penelitian ini menerapkan pembobotan kata atau TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF merupakan suatu metode pembobotan kata terhadap suatu dokumen. Metode ini menggabungkan dua konsep perhitungan bobot yaitu frekuensi kemunculan kata dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan dari kata di dalam dokumen yang diberikan, akan menunjukkan seberapa pentingnya kata di dalam sebuah dokumen tersebut. Frekuensi dari dokumen yang memiliki kata tersebut, akan menunjukkan seberapa umumnya jenis kata tersebut. Bobot kata akan semakin besar, apabila sering muncul didalam sebuah dokumen, dan akan semakin kecil jika kemunculannya berada di dalam banyak dokumen[16]. Pada Algoritma TF-IDF digunakan rumus dengan persamaan (1) sebagai berikut:

$$IDF = \log \frac{N}{df} \quad (1)$$

Keterangan

IDF = *Inverse Document Frequency*

N = jumlah banyak dokumen/data

df = banyak dokumen yang mengandung kata yang dicari

2.5 Klasifikasi Improved KNN

Penelitian ini menggunakan metode *Improved KNN*. *Improved K-Nearest Neighbor* merupakan pengembangan dari metode *K-Nearest Neighbor* (K-NN), yang mana perbedaannya terdapat pada pembobotan nilai-k. Pada *Improved KNN* dilakukan perhitungan bobot pada tiap data latih berdasarkan data uji dan data latih lainnya, serta kelas dari data uji ditentukan oleh bobot data latih yang terbesar[17]. Untuk menghitung kemiripan antara dua dokumen, digunakan metode *Cosine Similarity* (CosSim). Rumus yang digunakan untuk menghitung *cosine similarity* dapat dilihat pada persamaan (2) sebagai berikut:

$$\cos(\theta_{QD}) = \frac{-b \pm \sqrt{b^2 - 4ac}}{\sqrt{\sum_{i=1}^n (Q_i)^2} \cdot \sqrt{\sum_{i=1}^n (D_i)^2}} \quad (2)$$

Keterangan:

$\cos(\theta_{QD})$ = Kemiripan dokumen Q terhadap dokumen D

Q = Data Uji

D = Data Latih

N = Banyaknya data

Persamaan di atas dilakukan untuk melakukan perhitungan penetapan *k-values* pada algoritma *Improved K-Nearest Neighbor*, dengan mengurutkan secara menurun terlebih dahulu hasil dari perhitungan similaritas pada setiap kategori.

3. HASIL DAN PEMBAHASAN

3.1 Proses Pengumpulan Data

Metode pengumpulan data dilakukan dengan dua teknik yaitu *crawling data* dari media sosial *twitter* dan *survey* kuesioner dari pengguna jasa LRT Kota Palembang. *Crawling data* adalah teknik pengumpulan data dari media sosial dengan menyertakan API (*Application Programming Interface*). *Twitter* API terdiri atas *Twitter Search* dan *Twitter Streaming* API. Keduanya memiliki perbedaan dimana *Twitter Search* API memiliki fungsi



pencarian ke masa lampau sedangkan untuk *Twitter Streaming API* memiliki fungsi pencarian ke masa yang akan datang[18]. Pada proses ini peneliti menggunakan *library python 3* untuk menarik data sentimen mengenai #LRTPalembang atau #LRTSumsel. Adapun *library* yang digunakan adalah *tweepy* dan *csv*, *library tweepy* merupakan *tools* bantu dengan menggunakan *access token* dan *access key* untuk menghubungkan data dari *twitter* kedalam *notebook jupyterlab* pada *python 3*. Lalu fungsi *library csv* merupakan proses untuk menyimpan data *tweet* tersebut kedalam file ekstensi *csv*. Adapun teknik pengumpulan data yang kedua yaitu survey kuesioner yaitu tahapan proses pengumpulan data dengan menyebarkan seperangkat pertanyaan kepada responden pengguna LRT Palembang. Pengumpulan data dilakukan melalui kuesioner yang berupa pendapat masyarakat langsung sebagai pengguna pelayanan dan fasilitas yang terdapat pada LRT Palembang. Adapun berikut gambar sampel data dengan teknik *crawling*. Pada **Tabel 1** merupakan sampel data yang diambil dengan cara *crawling data* dari media sosial *twitter*.

Tabel 1. Sampel Data *Crawling Twitter*

text
b'Akhirnya naik LRT di Palembang\nhttps://t.co/dzpa68vCYn\n\n#lrtpalembang'
b'Bagi turis yang main ke Palembang mau coba naik lrt, setelah tap liat depart keretanya, klo keretanya 5 menit nyampe boleh naik ke peron klo blm jgn naik ke peron krm gak diperbolehkan'
b'felt so lucky di Palembang ada lrt , jadi remaja remaja depresi seperti kita dapat menggunakannya untuk stress relief and go for short trip xixi https://t.co/GHSbSdt18Q'
b'KAI Palembang tambah delapan perjalanan LRT Sumsel https://t.co/lgAtKvnjSv'
b'Akhirnya naik LRT di Palembang\nhttps://t.co/dzpa68vCYn\n\n#lrtpalembang'

3.2 Preprocessing Text

Pada tahap preprocessing terdapat beberapa tahapan lainnya yang akan dilakukan, dimulai dari *case folding*, *tokenizing*, *filtering* dan *stemming*. Pada tahapan ini akan dimasukan data berupa teks yang berasal dari data yang sudah dikumpulkan mengenai LRT Kota Palembang. Sebelum melakukan proses klasifikasi, pada penelitian ini proses yang dilakukan adalah melakukan proses pemanggilan operator serta mengakses file data sentimen yang sudah di *sorting* yang memiliki format CSV atau sudah dinamakan "data.csv" serta sudah dikumpulkan dan akan diolah kedalam tahapan *preprocessing*. Pada proses ini peneliti menggunakan *library pandas* dan *numpy* untuk memanggil data yang telah dikumpulkan. Adapun tahapan *preprocessing*:

a. Case Folding

Tahapan ini merupakan proses penyamaan bentuk huruf pada dokumen teks. Proses yang dilakukan yaitu perubahan kata-kata yang sebelumnya mengandung huruf "*capital*" menjadi huruf "*lowercase*". proses pertama dari rangkaian *preprocessing* ialah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf A-Z yang diterima, penyamaan huruf dilakukan karena perbedaan bentuk huruf pada data sentimen akan mengganggu proses selanjutnya[19]. Proses ini merupakan pemanggilan operator *transform cases* yang dimana proses dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil semua atau menjadi huruf kapital semua,

Tabel 2. Data Proses *Case Folding*

Input	Output
Untuk Pelayanan Sudah Cukup Memuaskan Tetapi Untuk Petugas Nya Kurang Ramah Untuk Fasilitas Sudah Bagus Hanya Saja Tempat Sholat Ada Beberapa Yang Kurang Bersih	untuk pelayanan sudah cukup memuaskan tetapi untuk petugas nya kurang ramah untuk fasilitas sudah bagus hanya saja tempat sholat ada beberapa yang kurang bersih

b. Tokenizing

Tahapan ini dilakukan proses pemotongan tiap kata dalam kalimat atau *parsing* dengan menggunakan spasi sebagai delimiter yang akan menghasilkan token berupa kata untuk memudahkan dalam proses mendapatkan hasil frekuensi kata tiap kalimat[20].

Tabel 3. Data Proses *Tokenizing*

Input	Output
untuk pelayanan sudah cukup memuaskan tetapi untuk petugas nya kurang ramah untuk fasilitas sudah bagus hanya saja tempat sholat ada beberapa yang kurang bersih	'untuk', 'pelayanan', 'sudah', 'cukup', 'memuaskan', 'tetapi', 'untuk', 'petugas', 'nya', 'kurang', 'ramah', 'untuk', 'fasilitas', 'sudah', 'bagus', 'hanya', 'saja', 'tempat', 'sholat', 'ada', 'beberapa', 'yang', 'kurang', 'bersih'

c. Filtering



Proses Filtering digunakan untuk mengambil kata-kata yang penting dari hasil tokenizing. Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan *stopword*. Misalnya penggunaan kata penghubung seperti dan, yang, serta, setelah, dan lainnya. Penghilangan *stopword* dilakukan karena dapat mengurangi ukuran *index*, level *noise* dan waktu pemrosesan[21].

Tabel 4. Data Proses Filtering

Input	Output
'untuk', 'pelayanan', 'sudah', 'cukup', 'pelayanan', 'memuaskan', 'petugas', 'ramah', 'memuaskan', 'tetapi', 'untuk', 'petugas', 'nya', 'fasilitas', 'bagus', 'sholat', 'bersih'	
'kurang', 'ramah', 'untuk', 'fasilitas', 'sudah', 'bagus', 'hanya', 'saja', 'tempat', 'sholat', 'ada', 'beberapa', 'yang', 'kurang', 'bersih'	

Tujuan utama dalam penerapan proses *stopwords* ini adalah untuk mengurangi jumlah kata dalam sebuah dokumen yang nantinya akan berpengaruh dalam kecepatan dan performa *Natural Language Processing* (NLP).

d. Stemming

Proses *stemming* merupakan tahapan menghapus huruf yang menjadi kata imbuhan dalam sebuah kata yang ada dalam sentimen yang telah dilakukan proses *preprocessing* sebelumnya. Seperti kata “pelayanan” menjadi “layan”, kata “bekerja” menjadi “kerja” dan “sepertinya” menjadi “seperti”. Imbuhan yang dihapus yaitu awalan, akhiran maupun awalan dan akhiran sehingga menjadi kata dasar[22].

Tabel 5. Data Proses Stemming

Input	Output
'untuk', 'pelayanan', 'sudah', 'cukup', 'memuaskan', 'pelayanan', 'memuaskan', 'petugas', 'ramah', 'tetapi', 'untuk', 'petugas', 'nya', 'kurang', 'ramah', 'fasilitas', 'bagus', 'sholat', 'bersih'	
'untuk', 'fasilitas', 'sudah', 'bagus', 'hanya', 'saja', 'tempat', 'sholat', 'ada', 'beberapa', 'yang', 'kurang', 'bersih'	

3.3 Pembobotan Kata TF-IDF

Setelah proses *preprocessing*, selanjutnya peneliti melanjutkan tahapan TF-IDF yang dimana dilakukan perhitungan bobot frekuensi kemunculan dari kata didalam dokumen yang diberikan. Hasil yang didapatkan setelah dilakukan pembobotan dapat dilihat pada Gambar 3 sebagai berikut:

term	rank
guna	42.316198
lrt	40.837990
stasiun	39.459010
baik	39.208800
sangat	39.197950
dan	39.028726
dengan	38.816884
sudah	38.405338
ruang	37.883166
untuk	37.748638
di	36.035226
tiket	35.341463
tumpang	34.365423

Gambar 3. Hasil Ranking Kata TF-IDF

Dari Gambar 3 dapat diketahui bahwa 3 kata yang paling banyak digunakan dalam sentiment masyarakat terhadap LRT Kota Palembang adalah kata “guna”, “lrt”, dan “stasiun”.

3.4 Analisis Sentimen

Selanjutnya, data sentimen yang sudah diproses pada tahap *preprocessing* dan Tf-IDF, sentimen dibagi menjadi 3 kelompok sentimen yaitu sentimen positif, sentimen negatif dan sentimen netral melalui proses pelabelan data



dengan menggunakan *Library Polarity*. Dengan menentukan *polarity* dari sentimen, dengan ketentuan jika sentimen memiliki *polarity* 0 maka memiliki sentimen netral, jika sentimen memiliki *polarity* > 0.0 maka memiliki sentimen positif, sebaliknya jika sentimen memiliki *polarity* < 0 maka memiliki sentimen negatif. Adapun grafik persentase dari pembagian data sentimen pada **Tabel 6**.

Tabel 6. Pembagian Data Sentimen

No	Sentimen	Jumlah
1	Positif	1107
2	Netral	205
3	Negatif	305
Total Data		1617

Dari tabel 5 dapat dilihat Sentimen terhadap LRT Kota Palembang yang berisi kalimat positif memiliki jumlah label data sebanyak 1107 atau 68%, lalu sentimen negatif berjumlah 305 atau 19%, serta sentimen netral yang berjumlah 205 atau 13% dari total seluruh sentimen yang didapatkan. Hasil tersebut membuktikan sebanyak 68% masyarakat memberikan sentimen positif terhadap Pelayanan dan Fasilitas dari LRT Kota Palembang.

3.5 Proses Klasifikasi Improved KNN

Pada proses ini merupakan proses utama dari sentimen analisis terhadap fasilitas dan layanan penggunaan LRT Palembang. Pada proses ini dilakukan uji coba analisis sentimen mengenai fasilitas dan layanan penggunaan LRT Palembang yang didapat dari Twitter serta melakukan penyebaran kuesioner menggunakan metode *Improved K-Nearest Neighbor* dengan menggunakan *tools python 3*, 1617 *record* data, dan dengan melakukan 5 kali uji coba *confusion matrix* yang hasil data ujinya dapat dilihat pada **Tabel 6**.

Tabel 7. Hasil Data Uji *Confusion Matrix*

No	Record Data	Data Training	Data Testing	Pembagian Data	Hasil Akurasi	Hasil Persentase
1	1617	1455	162	90:10	0.7407	74.07%
2	1617	1294	323	80:20	0.6882	68.82%
3	1617	1132	485	70:30	0.6810	68.10%
4	1617	970	647	60:40	0.6445	64.45%
5	1617	809	809	50:50	0.6304	63.04%

Berdasarkan hasil uji coba pada **Tabel 7** untuk mencari nilai akurasi dengan 1617 *record data* dan partisi persentase yang berbeda-beda sebanyak 5 kali maka memiliki hasil tertinggi pada pengujian data 90:10 yang dimana hasil akurasi nya sebesar 0.7407 atau memiliki hasil sebesar 74.07 %, sedangkan hasil terendah pada pengujian data 50.50 yang dimana memiliki hasil 0.6304 atau 63.04%.

	precision	recall	f1-score	support
0	0.59	0.53	0.56	36
1	0.71	0.25	0.37	20
2	0.78	0.91	0.84	106
accuracy			0.74	162
macro avg	0.70	0.56	0.59	162
weighted avg	0.73	0.74	0.72	162

Gambar 4. Hasil *Recall* dan *Precision*

Dari **Gambar 4** dapat dilihat dengan nilai akurasi 74.07% yang didapat dari pembagian data 90:10 memiliki nilai rata-rata presisi 70%, rata-rata recall 56% dan rata-rata nilai f-1 score 59% dengan pengujian menggunakan *python 3* pada analisis sentiment masyarakat terhadap LRT Kota Palembang.

4. KESIMPULAN

Penelitian analisis sentimen ini menerapkan metode *Improved K-Nearest Neighbor* yang merupakan modifikasi dari metode *K-Nearest Neighbor*, metode ini dipilih karena efektif digunakan dalam jumlah data yang besar dan juga tangguh terhadap data *noisy*. Perbedaan metode ini dengan metode sebelumnya terletak pada penentuan nilai-k, dimana penentuan nilai-k tetap dilakukan hanya saja setiap kategori memiliki nilai-k yang berbeda. Data yang digunakan pada penelitian ini berjumlah 1617 *record* data sentimen yang didapatkan dari proses dua teknik pengumpulan data yaitu, *crawling data twitter* dan survey kuesioner pengguna LRT Palembang. *Tweets* dan



pendapat yang digunakan pada penelitian ini mengandung *keyword* “LRT Palembang” atau “LRT Sumsel”. Proses analisis *document* dalam penelitian ini menggunakan perangkat lunak *python 3* yang dimana *sentiment* dikelompokkan menjadi 3 kategori yaitu *sentiment positif*, *sentiment negatif* dan *netral*. Berdasarkan penelitian yang telah dilakukan, didapatkan hasil *sentimen masyarakat terhadap LRT Kota Palembang* dengan menggunakan metode *Improved K-Nearest Neighbor* dengan hasil *sentimen positif 67%*, *sentimen negatif 19%* dan *sentimen netral 13%* dan dengan penerapan metode *Improved K-Nearest Neighbor* untuk analisis *sentimen masyarakat terhadap LRT Kota Palembang*, menghasilkan nilai akurasi tertinggi 74.07% pada pembagian data testing dan data training 90:10 dengan nilai presisi 70%, recall 56% dan f-1 score 59%, sedangkan nilai akurasi terendah didapatkan pada pembagian data testing dan data training 50:50 dengan nilai akurasi terendah 63.04%. Hasil penelitian ini diperoleh dari analisis *sentimen dari masyarakat terhadap LRT kota Palembang* yang dilakukan dengan bantuan *tools python 3* yang diimplementasikan menggunakan *ipython lab*.

REFERENCES

- [1] R. S. Salim, “Perubahan Beban Di Feeder Dc Switchgear Lrt Palembang,” 2021, [Online]. Available: <http://repository.univ-tridnanti.ac.id/2985/>.
- [2] D. Widiyanti, “Pengembangan Park and Ride untuk Meningkatkan Pelayanan Angkutan LRT Kota Palembang,” *Jurnal Penelitian Transportasi Darat.*, vol. 21, pp. 103–116, 2019, doi:10.25104/jptd.v21i2.1562.
- [3] W. M. Baihaqi et al., “Kombinasi K-Means Dan Support Vector Machine (Svm) Untuk K-Means And Support Vector Machine (Svm) Combination To Predict Sara Elements On Tweet,” vol. 7, no. 3, pp. 501–510, 2020, doi: 10.25126/jtiik.202072126.
- [4] W. Oktinas, “Analisis Sentimen Pada Acara Televisi menggunakan Improved K-Nearest Neighbor,” *Program di Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan*, vol.1, no.2, pp. 6-38, 2017.
- [5] G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, “Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Able Propagation,” *CITEE*, pp.7-8, 2015.
- [6] B. Li and L. Han, “Distance weighted cosine similarity measure for text classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8206 LNCS, pp. 611–618, 2013, doi: 10.1007/978-3-642-41278-3_74.
- [7] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification,” *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE* 197, vol. 2018-Janua, pp. 294–298, 2018, doi: 10.1109/ICITISEE.2017.8285514.
- [8] N. Octaviani Faomasi Daeli, “Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor,” *Open Access J Data Sci Appl*, vol. 3, no. 1, pp. 1–007, 2020, doi: 10.34818/JDSA.2020.3.22.
- [9] A. Salam, J. Zeniarja, and R. S. U. Khasanah, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.
- [10] 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, doi: 10.14421/jiska.2018.31-01.
- [11] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JA 20 (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [12] A. Nurzahputra, M. A. Muslim, and M. Khusniati, “Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa,” *Techno.Com*, vol. 16, no. 1, pp. 17–24, 2017, doi: 10.33633/te.v16i1.1284.
- [13] C. Juditha, “Sentimen Dan Imparsialitas Isi Berita Tentang Ahok Di Portal Berita Online,” *J. Penelit. Komun. dan Informatika*, vol. 18, no. 1, p. 57, 2017, doi: 10.31346/jpkp.v18i1.839.
- [14] N. L. Ratniasih, M. Sudarma, and N. Gunantara, “Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat,” *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 3, p. 13, 2017, doi: 10.24843/mite.2017.v16i03p03.
- [15] G. Divva, M. Zulma, and N. Chamidah, “Perbandingan Metode Klasifikasi Naive Bayes , Decision Tree Dan K-Nearest Neighbor Pada Data Log Firewall,” no. April, pp. 679–688, 2021.
- [16] K. A. Ghofari, N. F. R. and L. Selmakarmy, “Pembuatan Sistem Pencarian Hadis dengan menggunakan Metode Pembobotan TF-IDF,” *Seminar Nasional Teknik Elektro, Sistem Informasi dan Teknik Informatika*, volume 1, pp. 207–212, 2021.
- [17] K. Telaumbanua, S. Sudarto, F. Butar-Butar, and P. S. Bilqis, “Identifikasi Sampah Berdasarkan Tekstur Dengan Metode GLCM dan GLRLM Menggunakan Improved KNN,” *Explorer (Hayward)*, vol. 1, no. 2, pp. 45–52, 2021, doi: 10.30865/explorer.v1i2.94.
- [18] D. Rustiana and N. Rahayu, “Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naive Bayes,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 8, no. 1, pp. 113–120, 2017, doi: 10.24176/simet.v8i1.841.
- [19] M. Z. Naf’an, A. Burhanuddin, and A. Riyani, “Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen,” *J. Linguist. Komputasional*, vol. 2, no. 1, pp. 23–27, 2019.
- [20] R. R. A. Siregar, F. A. Sinaga, and R. Arianto, “Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 1, no. 2, p. 171, 2017, doi: 10.24912/computatio.v1i2.1014.
- [21] A. Aziz, F. Fauziah, and I. Fitri, “Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 Menggunakan Metode Naive Bayes,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 5, no. 2, pp. 842–851, 2021, [Online]. Available: <https://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/381>.
- [22] A. Y. Pratama, Y. Umaidah, and A. Voutama, “Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja),” *Sains Komput. Inform.*, vol. 5, no. 2, pp.



897-910, 2021, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/386/365>.

Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor

ORIGINALITY REPORT

23%

SIMILARITY INDEX

21%

INTERNET SOURCES

14%

PUBLICATIONS

12%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Abdul Mizwar A. Rahim, Guido Adolfus Suni, Setefensius Sasi, Galang Cahya Pangestu, Maikel Fainsenem, Muhammad Arsyad A. "Implementasi Algoritma Apriori Untuk Mengetahui Pola Pembelian Di Starcomp Jogja", Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 2020
Publication 2%
- 2 Submitted to Universitas Nasional
Student Paper 2%
- 3 repository.unsri.ac.id
Internet Source 1%
- 4 Tiadi Widiyanto. "Rancang Bangun Aplikasi Chatbot Untuk Pendukung Perdagangan Dengan Menggunakan Metode Fuzzy String Matching-RUP (Studi Kasus : Warung Kedelai Edamame Kalibagor)", EVOLUSI : Jurnal Sains dan Manajemen, 2022
Publication 1%

5	jurnal.usbypkp.ac.id Internet Source	1 %
6	ejournal.undiksha.ac.id Internet Source	1 %
7	www.grafiati.com Internet Source	1 %
8	id.scribd.com Internet Source	1 %
9	www.jurnal.iaii.or.id Internet Source	1 %
10	ejurnal.itats.ac.id Internet Source	1 %
11	journal.budiluhur.ac.id Internet Source	1 %
12	Submitted to Universitas Trunojoyo Student Paper	1 %
13	senafti.budiluhur.ac.id Internet Source	1 %
14	Kamal Thennakoon, Dasun Ekanayake, Thanura Marapana, Ashen Ranasinghe, Dinuka R. Wijendra, Anjalie Gamage. "PROBEXPERT: An Enhanced Q&A Platform for Reducing Time Spent on Learning and Finding Answers", 2022 IEEE 7th International	1 %

conference for Convergence in Technology (I2CT), 2022

Publication

15	Submitted to Universitas Sanata Dharma Student Paper	1 %
16	ojsbimtek.univrab.ac.id Internet Source	1 %
17	repository.nusamandiri.ac.id Internet Source	1 %
18	jurnal.fikom.umi.ac.id Internet Source	1 %
19	shmpublisher.com Internet Source	1 %
20	ejournal.uksw.edu Internet Source	1 %
21	ejournal.upbatam.ac.id Internet Source	1 %
22	ejournal.uin-suka.ac.id Internet Source	1 %
23	notarazzi.blogspot.com Internet Source	1 %
24	ecampus.pelitabangsa.ac.id Internet Source	1 %
25	Submitted to Chungnam National University Student Paper	1 %

26

e-journal.uajy.ac.id

Internet Source

1 %

Exclude quotes Off

Exclude matches < 1%

Exclude bibliography Off