

Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor

Restu Ramadhan*, M Afdal, Inggih Permana, Muhammad Jazman

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: ^{1,*}restu301297@gmail.com, ²m.afdal@uin-suska.ac.id, ³inggihpermana@uin-suska.ac.id, ⁴jazman@uin-suska.ac.id
Email Penulis Korespondensi: restu301297@gmail.com

Submitted 18-06-2023; Accepted 30-06-2023; Published 30-06-2023

Abstrak

Transportasi online adalah Inovasi dalam teknologi yang muncul untuk memecahkan berbagai masalah yang timbul pada transportasi umum konvensional seperti halnya dalam kemudahan pemesanan, ketersediaan dan digitalisasi pembayaran. Maxim adalah perusahaan transportasi online yang sudah beroperasi sejak tahun 2018 di Indonesia. Seiring meningkatnya jumlah pengguna aplikasi maxim, membuat tuntutan terhadap mutu pelayanan aplikasi juga meningkat. Dalam Google Play Store, ulasan dan informasi suatu aplikasi disimpan dalam bentuk teks. Salah satu proses pengambilan informasi text mining dalam kategori teks adalah Analisis Sentimen untuk melihat kecenderungan suatu sentimen atau pendapat apakah bersifat positif, netral, atau negatif pada ulasan pengguna aplikasi Maxim. Proses klasifikasi sentimen menggunakan algoritma K-NN menghasilkan akurasi, presisi, dan recall berturut-turut sebesar 90,23%; 90,23%; dan nilai recall 72,38% dengan percobaan menggunakan 90% data latihan dan 10% data uji serta nilai $k = 5$.

Kata Kunci: Analisis Sentimen; Google Play Store; K-Nearest Neighbor; Maxim; Text Mining

Abstract

Online transportation is an innovation in emerging technology to solve various problems that arise in conventional public transportation such as in the ease of ordering, availability, and digitization of payments. Maxim is an online transportation company that has been operating since 2018 in Indonesia. As the number of users of the maxim application increases, demands for the quality of application service also increase. In the Google Play Store, reviews and information about an app are stored in text form. One of the processes of extracting text mining information in the text category is Sentiment Analysis to see the tendency of a sentiment or opinion whether it is positive, neutral, or negative at the Maxim application user reviews. The sentiment classification process using the K-NN algorithm produces accuracy, precision, and recall of 90.23%; 90.23%; and a recall value of 72.38% with an experiment using 90% training data, 10% test data, and a value of $k = 5$.

Keywords: Google Play Store; K-Nearest Neighbor; Maxim; Sentiment Analysis; Text Mining

1. PENDAHULUAN

Transportasi online adalah suatu penyelenggaraan Lalu Lintas dan Angkutan Jalan yang berjalan dengan mengikuti serta memanfaatkan perkembangan ilmu pengetahuan (teknologi) berbasis aplikasi dan online baik untuk pemesanan maupun pembayaran [1]. Inovasi ini muncul untuk memecahkan berbagai masalah yang timbul pada transportasi umum konvensional seperti kemudahan pemesanan, ketersediaan dan digitalisasi pembayaran. Perkembangan teknologi yang sangat pesat membawa pengaruh yang besar bagi manusia. Salah satu pengaruh dari perkembangan teknologi adalah semakin maraknya penggunaan transportasi online melalui media internet. Menurut data dari riset *Google*, Temasek dan Bain & Company [2]. Pertumbuhan angka pasar transportasi online di Indonesia mengalami pertumbuhan sangat besar, pada tahun 2015 bernilai US\$ 900 juta meningkat di tahun 2018 sebesar US\$ 12,7 miliar dan di prediksi mencapai US\$ 40 miliar pada 2025. Dari riset tersebut juga menunjukkan bahwa Indonesia merupakan negara dengan pertumbuhan pangsa pasar transportasi online terbesar di Asia Tenggara. Sementara itu, berdasarkan survei yang dilakukan Asosiasi Penyelenggara Jasa Internet Indonesia memaparkan pemesanan transportasi online menempati urutan ke 16 dari 22 alasan mengapa masyarakat menggunakan internet dan 93,9% masyarakat mengakses internet menggunakan smartphone [3].

Terdapat beberapa transportasi online di Indonesia yaitu, Gojek, Grab, Maxim, FastGo, Bitcar, Bonceng, Anterin.id [4]. Maxim adalah perusahaan internasional yang bergerak dibidang teknologi informasi dimana perusahaan ini menyediakan platform yang mempertemukan pengemudi dan pelanggan secara mudah. Maxim adalah perusahaan transportasi online yang bisa dibilang baru di Indonesia. Maxim sendiri sudah ada sejak tahun 2003 lalu. Namun dulunya hanya sebuah layanan aplikasi taksi di kota Chardinsk yang terletak di pegunungan Ural, Rusia. Di Indonesia maxim sendiri pertama kali beroperasi pada tahun 2018. Perusahaannya semakin melebarkan sayap dengan tidak hanya menjadi perusahaan transportasi online yang berfokus ke taksi saja, melainkan juga jenis layanan angkutan lain seperti ojek atau mobil pada umumnya. Maxim mengembangkan sendiri aplikasinya dan menciptakan sistem perangkat keras dan perangkat lunak yang memungkinkan mitra-mitranya terhubung kelayanan perusahaan, memperoses jutaan setiap hari, memantau kualitas kerja dan pelayanan, serta menganalisis dan mengoptimalkan bisnis mereka. Maxim memiliki kebijakan yang menguntungkan bagi pengemudi maupun pelanggan, seperti jadwal yang fleksibel untuk pengemudi, harga terjangkau serta sistem reservasi order untuk pelanggan [5].

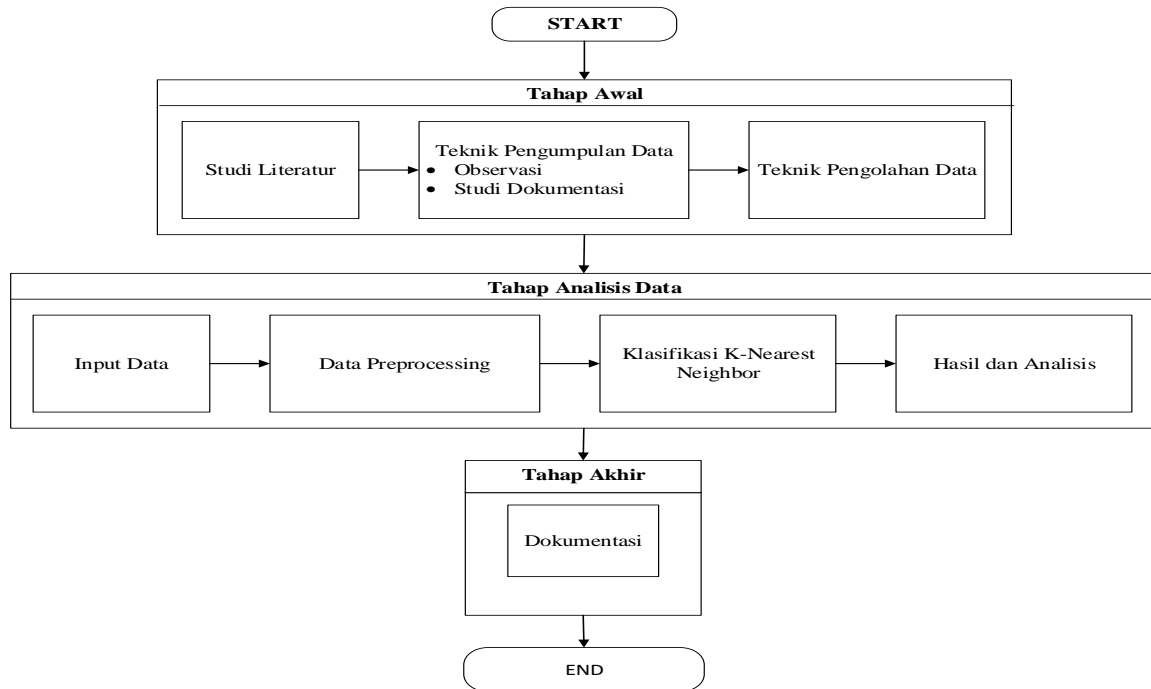
Seiring meningkatnya jumlah pengguna *smartphone* dan kendaraan berakibat pada semakin banyaknya para pengguna aplikasi maxim, hal ini membuat tuntutan terhadap mutu pelayanan aplikasi juga meningkat. Dalam Google Playstore setiap pengguna dapat memberikan rating dan ulasan terhadap suatu aplikasi. Menurut [6], ulasan dan informasi suatu produk disimpan dalam bentuk teks, maka text mining merupakan solusi dalam pengambilan informasi yang berbentuk teks. Salah satu proses pengambilan informasi text mining dalam kategori teks adalah *Sentiment Analysis* atau

Analisis Sentimen. Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks [7]. Analisis sentimen diperlukan sebagai alat untuk mengklasifikasikan sebuah informasi yang berbentuk teks bahasa kedalam kategori positif dan negatif [8]. Analisis sentimen digunakan untuk melihat kecenderungan suatu sentimen atau pendapat, apakah pendapat tersebut cenderung beropini positif atau negatif. Menurut [9], analisis sentimen pada review pengguna adalah proses menyelidiki review produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan. Analisis sentimen adalah teknik komputasi pendapat, perasaan dan subjektivitas teks [10]. Tugas dasar dalam analisis sentimen adalah untuk mengklasifikasikan polaritas teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas berarti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif. Oleh karena itu, dibutuhkan suatu analisis terhadap opini-opini tersebut dalam penelitian ini agar bisa dijadikan tolak ukur baik atau tidaknya jasa transportasi online menurut pelanggan. Tujuan dari analisis sentimen selain untuk mendapatkan informasi sentimen yang terkandung dalam sebuah teks, juga untuk melihat pendapat teks yang berkaitan dengan masalah atau objek, apakah itu cenderung memiliki pandangan positif atau negatif.

Proses pengklasifikasian opini melalui pendekatan text mining memerlukan metode yang mampu mengklasifikasikan opini secara akurat. Penelitian sebelumnya terkait analisis sentimen pernah dilakukan untuk menganalisis opini mahasiswa terhadap penilaian kinerja dosen berdasarkan kuesioner dengan menerapkan algoritma *K-Nearest Neighbor* (KNN) yang menghasilkan akurasi sebesar 81%, *precision* sebesar 81%, *recall* sebesar 81%, dan *F1-Score* sebesar 81% [11]. Penelitian lainnya pada tahun 2020 dilakukan oleh Angelina, *dkk* untuk menganalisis sentimen pengguna Twitter terhadap Aplikasi Ruang Guru yang menghasilkan nilai optimal pada model optimasi algoritma PSO berbasis SVM dengan nilai akurasi sebesar 78,55% dan AUC sebesar 0,853 [12]. Sehingga dalam pengklasifikasian ini diterapkan algoritma KNN dan pembobotan untuk setiap kata menggunakan TF-IDF. Ada beberapa kelebihan dari metode K-NN yaitu algoritma klasifikasi *K-Nearest Neighbor* yang terbukti mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan dalam sebuah aplikasi [13]. Hal tersebut diketahui karena algoritmanya yang mudah dan efisien untuk klasifikasi teks. Bukan hanya mudah dan efisien, sifat dari algoritma *K-Nearest Neighbor* sendiri bersifat *self-learning*, dimana algoritma ini dapat mempelajari struktur data yang ada dan mengkategorikan dirinya sendiri [14]. Penelitian ini hanya berfokus menganalisa kepuasan pengguna atau pelanggan Maxim, dimana kepuasan adalah suatu sikap yang diputuskan berdasarkan pengalaman yang didapat [15]. Penelitian ini menggunakan data ulasan pengguna Maxim untuk kemudian diproses dan dianalisis dengan algoritma KNN sehingga menghasilkan klasifikasi sentimen terhadap tiap ulasan pengguna yang dibagi menjadi tiga kelas sentimen yaitu sentimen positif, negatif, dan netral.

2. METODOLOGI PENELITIAN

Tahapan penelitian ini digambarkan melalui diagram alir metodologi penelitian pada Gambar 1. Tahap awal dalam penelitian adalah melakukan studi literatur untuk menemukan landasan teori dan referensi tertulis yang mendasari penelitian guna menyelesaikan permasalahan serta dijadikan referensi yang kuat dalam melakukan analisa. Pemrosesan data dimulai dengan (a) Tahap praproses data; (b) Pembagian data latih dan data uji; (c) Pemodelan klasifikasi data dengan K-NN; dan (d) Analisis sentimen hasil klasifikasi. Tahapan praproses data ulasan terdiri dari tahap *case folding*, *tokenizing*, *cleansing*, *stopword removal*, *stemming*, pebelan data ulasan, dan diakhiri dengan proses pembobotan TF-IDF. Setelah data selesai diproses kemudian dilakukan proses distribusi data latih dan data uji dengan kuantitas yang beragam, kemudian dilakukan pemodelan data dengan KNN menggunakan beberapa nilai parameter *k*. Data yang digunakan merupakan data yang diperoleh dari ulasan pengguna aplikasi Maxim selama satu bulan terhitung dari tanggal 7 Agustus 2022 sampai dengan 7 September 2022 dengan jumlah 10291 ulasan. Data ulasan pengguna aplikasi Maxim diperoleh di *Google Play Store* yang dapat diakses melalui tautan <https://play.google.com/store/apps/details?id=com.taxsee.taxsee&hl=id&gl=US&pli=1> dengan menggunakan teknik *scraping* melalui bahasa pemrograman *Python*. Proses *scraping* data ulasan menggunakan *Package google_play_scraper* yang tersedia di *python*.



Gambar 1. Alir Diagram Penelitian

2.1 Analisis Sentimen

Analisis sentimen adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial berdasarkan fakta dan opini [16]. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut, dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien. Sentiment analysis merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks.

2.2 Text Mining

Text mining adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data. Ini juga dikenal sebagai analisis teks, meskipun beberapa orang menarik perbedaan antara dua istilah; dalam pandangan itu, analitik teks adalah aplikasi yang diaktifkan oleh penggunaan teknik *text mining* untuk memilah-milah set data [17] [18].

2.3 Text Preprocessing

Berdasarkan ketidakteraturan struktur data teks, maka proses sistem temu kembali informasi ataupun *text mining* memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Salah satu implementasi dari *text mining* adalah tahap *text preprocessing* [19]. Tahap *Preprocessing* diantaranya:

a. Case Folding

Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau lowercase). *Case folding* adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf „a“ sampai dengan „z“ yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter*.

b. Tokenizing

Tahap *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

c. Filtering

Tahap Filtering adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). Stoplist/stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya.

d. Stemming

Teknik Stemming diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan katakata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

2.4 K- Nearest Neighbour (KNN)

Algoritma *k-Nearest Neighbor* adalah algoritma supervised learning dimana hasil dari instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori *k*-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sample-sample dari training data. Algoritma *k-Nearest Neighbor* menggunakan *Neighborhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru [20].

3. HASIL DAN PEMBAHASAN

Data ulasan pada Tabel 1 yang diperoleh dari hasil *scraping* tidak dapat langsung digunakan pada proses mining dengan algoritma Data Mining KNN. Data ulasan perlu dipraproses untuk menghilangkan data noise, *missing values*, atau data ulasan yang tidak memiliki informasi yang berguna, tidak memiliki makna, dan data yang sekiranya dapat mengganggu proses mining.

Tabel 1. Ulasan Pengguna Aplikasi Maxim

No	Ulasan	Tanggal
1	Lokasinya tambahin dong	07 September 2022
2	drvr nya ramah	07 September 2022
3	sangan bagus dan baik	07 September 2022
4	makasih ya driver	07 September 2022
5	Ok trmks	07 September 2022
6	sangat membantu Dan sopir nya sang at ramah	07 September 2022
7	Tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah	07 September 2022
8	sangat memuaskan	07 September 2022
...
10290	terimakasih banyak..supirnya sangat santun dan baik banget..	07 August 2022
10291	Bagusss	07 August 2022

3.1 Praproses Data

Tahapan praproses data ulasan terdiri dari tahap *case folding*, *tokenizing*, *cleansing*, *stopword removal*, *stemming*, pelebaban data ulasan, dan diakhiri dengan proses pembobotan TF-IDF.

3.1.1 Tahap Case Folding

Case folding merupakan tahap untuk mengubah ulasan yang berbentuk teks atau string berhuruf kapital atau berhuruf kapital dan huruf kecil, menjadi keseluruhan teks berhuruf kecil. Hal ini dilakukan agar setiap string yang memiliki susunan kata yang sama hanya akan membentuk satu fitur dari susunan kata tersebut. Berikut merupakan hasil dari proses *Case Folding* pada Tabel 2.

Tabel 2. Data Hasil Case Folding

No	Ulasan	Case Folding
1	Lokasinya tambahin dong	lokasinya tambahin dong
2	drvr nya ramah	drvr nya ramah
3	sangan bagus dan baik	sangan bagus dan baik
4	makasih ya driver	makasih ya driver
5	Ok trmks	ok trmks
6	sangat membantu Dan sopir nya sang at ramah	sangat membantu dan sopir nya sang at ramah
7	Tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah	tolong perbaiki masa sy nungguin sampe 20 menit y udah sy batalin aja cari yg sebelah
8	sangat memuaskan	sangat memuaskan
...
10290	terimakasih banyak..supirnya sangat santun dan baik banget..	terimakasih banyak..supirnya sangat santun dan baik banget..
10291	Bagusss	bagusss

3.1.2 Tahap Tokenizing

Tahap *tokenizing* merupakan proses memecahkan kalimat atau string menjadi token berupa potongan karakter atau kata. Contohnya seperti kalimat “titik jemput akurat” akan dipecah menjadi potongan-potongan kata “titik”, “jemput”, dan “akurat”. Proses *tokenizing* pada ulasan menggunakan *library nltk* yang tersedia di *python*. Berikut merupakan hasil pada data yang dilakukan proses *tokenizing* pada Tabel 3.

3.1.3 Tahap Cleansing

Tahap *Cleansing* dibagi menjadi dua tahap, tahap pertama adalah proses perubahan kata slang atau kata tidak baku pada ulasan, proses normalisasi mengacu pada kamus kata normalisasi yang digunakan untuk mengubah data tidak baku menjadi kata baku atau normal. Seperti kata “7an” dinormalisasi menjadi kata “tujuan”; kata “abis” menjadi “habis”; kata “adlah” menjadi “adalah”, dan lain sebagainya. Tahap selanjutnya adalah tahap *filtering* yang merupakan proses untuk menghapus karakter dalam string yang bukan bagian dari alfabet dan tidak memiliki makna tertentu sehingga mengurangi karakter yang tidak dibutuhkan, seperti angka; *hashtag*; emoji; *link*; tanda baca, dan lain sebagainya. Berikut merupakan hasil akhir dari proses *cleansing* data ulasan pada Tabel 3.

Tabel 3. Data Hasil *Tokenizing & Cleansing*

No	Tokenizing	Cleansing
1	['lokasinya', 'tambahin', 'dong']	lokasinya tambah dong
2	['drvr', 'nya', 'ramah']	drvr nya ramah
3	['sangan', 'bagus', 'dan', 'baik']	sangan bagus dan baik
4	['makasih', 'ya', 'driver']	terima kasih iya driver
5	['ok', 'trmks']	oke terima kasih
6	['sangat', 'membantu', 'dan', 'sopir', 'nya', 'sang', 'at', 'ramah']	sangat membantu dan sopir nya sang di ramah
7	['tolong', 'perbaiki', 'masa', 'sy', 'nungguin', 'sampe', 'menit', 'y', 'udah', 'sy', 'batalin', 'aja', 'cari', 'yg', 'sebelah']	tolong perbaiki masa saya menunggu sampai menit ya sudah saya batal saja cari yang sebelah
8	['sangat', 'memuaskan']	sangat memuaskan
...
102	['terimakasih', 'banyak', 'supirnya', 'sangat', 'santun', 'dan', 'baik', 'banget']	terima kasih banyak supirnya sangat santun dan baik banget
102	['bagusss']	bagus
91		

3.1.4 Tahap *Stopword Remove*

Tahap *stopword remove* adalah tahap untuk menghilangkan kata yang umum dan tidak memiliki arti tertentu, seperti kata konjungsi atau kata hubung dan kata adverbial atau kata keterangan yang tidak memiliki dampak pada tahap analisis mining. Seperti kata “saya”, “dia”, “bawah”, “dengan”, “nya”, “itu”, “ini”, dan kata lainnya. Tahap *stopword remove* menggunakan kamus *stopword* yang mengacu dengan *stoplist* pada penelitian [21] dan kata-kata lainnya merupakan hasil dari pengumpulan yang dilakukan oleh peneliti. Berikut merupakan hasil dari tahap *stopword remove* menyisihkan data akhir sebanyak 4269 *raw data* sebagaimana pada Tabel 4.

3.1.5 Tahap *Stemming*

Proses *stemming* pada penelitian ini menggunakan *library* Sastrawi yang tersedia di Python. Proses ini bertujuan untuk mengubah kata ke bentuk dasarnya sebagaimana aturan dalam Kamus Besar Bahasa Indonesia (KBBI), misalnya seperti kata “menunggu” menjadi kata dasar “tunggu”; kata membantu menjadi kata “bantu”; dan lain sebagainya. Adapun hasil dari tahap *stemming* tertera pada Tabel 4.

Tabel 4. Hasil *Stopword Remove* dan *Stemming*

No.	<i>Stopword Remove</i>	<i>Stemming</i>
1	lokasinya	lokasi
2	drvr ramah	drvr ramah
3	bagus baik	bagus baik
4	terima kasih driver	terima kasih driver
5	oke terima kasih	oke terima kasih
6	membantu sopir sang ramah	bantu sopir sang ramah
7	tolong perbaiki menunggu batal cari sebelah	tolong baik tunggu batal cari belah
8	memuaskan	muas
...
10290	terima kasih supirnya santun baik	terima kasih supirnya santun baik
10291	bagus	bagus

Berdasarkan hasil dari proses casefolding hingga proses stemming diperoleh 269 record data yang *missing values* atau *null*/kosong, hal ini disebabkan data tersebut tidak memiliki makna atau informasi berguna berdasarkan hasil praproses, sehingga data akhir yang dapat digunakan untuk proses selanjutnya berjumlah 10022 *record* data ulasan.

3.1.6 Tahap *Pelebelan*

Data ulasan merupakan data teks yang bersifat tidak terstruktur dan tidak memiliki kelas/label, agar dapat digunakan dalam proses klasifikasi dengan algoritma K-NN, maka perlu dilakukan proses pelebelan data. Proses pelebelan data memerlukan kamus sentimen kata positif dan kata negatif. Kamus kata negatif dan positif yang digunakan pada penelitian

ini adalah kamus berdasarkan rujukan penelitian [22] dan dilengkapi dengan kamus yang terdapat pada *repository* milik Ramaprakoso (2017) [23] serta berdasarkan tinjauan Pustaka oleh peneliti. Adapun tahapan dalam melakukan pelebelan data teks adalah sebagai berikut [24]:

- a. Menentukan term yang mengacu dan mampu merepresentasikan kata positif dan negatif, dalam hal ini adalah adanya kamus sentimen kata positif dan kata negatif. Kata positif adalah kata yang menggambarkan sanjungan, pujian, atau yang bermakna baik dan terpuji, sedangkan kata negatif merupakan kata yang merepresentasikan kata yang memiliki arti tidak atau kurang baik yang biasanya mengarah pada antonim atau negasi dari kata-kata pujian atau positif.
- b. Melakukan penghitungan jumlah kata positif dan negatif pada setiap kalimat, dimana ketika jumlah kata positif lebih besar dari jumlah kata negatif pada kalimat, maka kalimat tersebut masuk ke kelas sentimen positif, sebaliknya jika jumlah kata positif lebih kecil dari jumlah kata negatif pada kalimat, maka Kelas kalimat tersebut masuk ke kelas sentimen negatif. Adapun jika jumlah kata positif sama dengan jumlah kata negatif dalam sebuah kalimat, maka kalimat tersebut masuk ke dalam kelas sentimen netral.

Tabel 7. Data Hasil Pelebelan

No	Stemming	Count	Sentimen
1	lokasi	0	netral
2	drv ramah	1	positif
3	bagus baik	2	positif
4	terima kasih driver	2	positif
5	oke terima kasih	3	positif
6	bantu sopir sang ramah	2	positif
7	tolong baik tunggu batal cari belah	-1	negatif
8	muas	0	netral
...
10021	terima kasih supirnya santun baik	4	positif
10022	bagus	1	positif

3.1.7 Tahap Pembobotan TF-IDF

Data hasil *stemming* akan digunakan sebagai acuan untuk membuat *fitur* data. Untuk menghasilkan *fitur* perlu menghitung bobot pada setiap term dengan metode perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF akan menghasilkan nilai dari setiap kata yang sudah diekstrak sebelumnya. Untuk menghitung bobot pada tiap *fitur* data dengan *library scikit-learn* pada bahasa pemrograman python menggunakan modul *TfidfVectorizer* dan *CountVectorizer*. Hasil proses pembobotan menghasilkan potongan term atau kata yang berfungsi sebagai atribut sebanyak 3368 term sebagaimana yang tertera pada Tabel 8.

Tabel 8. Hasil Pembobotan TF-IDF

No	Term/Kata															
	abai	abal	abang	abg	abu	ac	aceh	acnya	acs	acuh	...	zendo	zone	zonk	zos	
	1	2	3	4	5	6	7	8	9	10	...	3365	3366	3367	3368	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
...	
6714	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	
6715	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	
...	
10021	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
10022	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

3.2 Pembagian Data Latih dan Data Uji

Untuk membuat pemodelan klasifikasi sentimen pada Data Ulasan, data yang telah melalui tahap pra-proses kemudian dibagi menjadi data latih dan data uji. Data latih adalah data yang digunakan untuk membuat model klasifikasi sedangkan data uji adalah data yang digunakan untuk menguji model yang telah dibangun dengan data latih. Adapun proporsi pembagian data pada penelitian ini menggunakan teknik pembagian data *Hold-out*. Proses pembagian data menggunakan modul *model_selection* yang terdapat pada *library scikit learn*, adapun untuk menentukan proporsi pembagian data yang paling optimal, dilakukan beberapa percobaan pembagian data sebagaimana tertera pada Tabel 9.

Tabel 9. Percobaan Pembagian Data

No.	Percobaan	Data Latih		Data Uji	
		Persentase	Jumlah Data	Persentase	Jumlah Data
1	P-1	50%	5011	50%	5011
2	P-2	60%	6013	40%	4009
3	P-3	70%	7015	30%	3007
4	P-4	75%	7516	25%	2506
5	P-5	80%	8017	20%	2005
6	P-6	85%	8518	15%	1504
7	P-7	90%	9019	10%	1003
8	P-8	95%	9620	5%	502

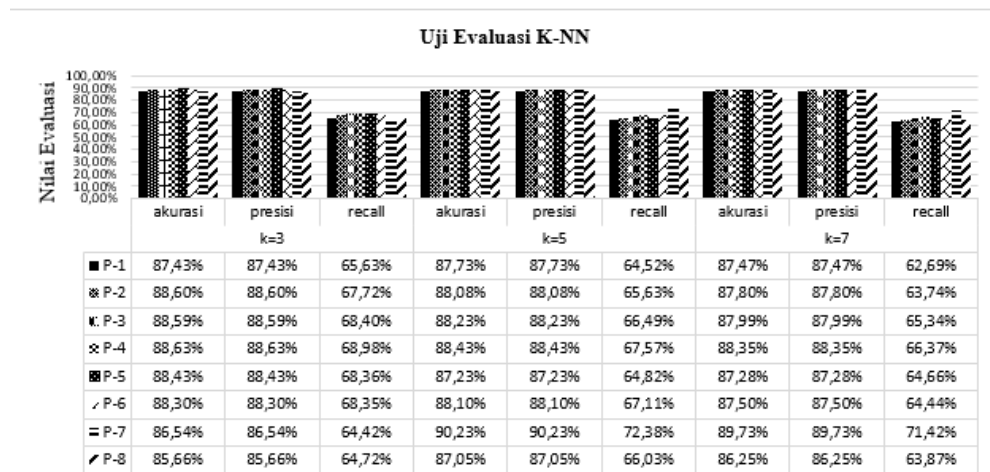
Karena sebaran sentimen ulasan yang tidak seimbang (*inbalanced*) dan kemungkinan memiliki pengaruh terhadap model klasifikasi, maka pemodelan akan dilakukan dengan menerapkan teknik rekonstruksi data SMOTE yang bertujuan untuk menyeimbangkan jumlah sentimen dengan meningkatkan jumlah sentimen minoritas melalui rekonstruksi ulang sentimen minoritas secara acak sehingga memiliki jumlah yang sama dengan sentimen mayoritas. Dalam hal ini kelas negatif dan netral akan memiliki jumlah yang sama dengan sentiment positif sebanyak 7978 ulasan. Sehingga secara keseluruhan penerapan SMOTE menghasilkan jumlah dataset sebanyak 23934 ulasan. Tabel 10 merupakan pembagian data latih dan data uji pada dataset hasil resampling SMOTE.

Tabel 10. Percobaan Pembagian Dataset *Resampling* SMOTE

No.	Percobaan	Data Latih		Data Uji	
		Persentase	Jumlah Data	Persentase	Jumlah Data
1	P-1	50%	11967	50%	11967
2	P-2	60%	14360	40%	9574
3	P-3	70%	16753	30%	7181
4	P-4	75%	17950	25%	5984
5	P-5	80%	19147	20%	4787
6	P-6	85%	20343	15%	3591
7	P-7	90%	21540	10%	2394
8	P-8	95%	22737	5%	1197

3.3 Pemodelan Sentimen dengan K-NN

Pemodelan klasifikasi sentimen dengan algoritma KNN menggunakan library *scikit learn* dengan modul *KNeighborsClassifier*. Adapun pemodelan pada penelitian ini menggunakan nilai parameter k atau jumlah banyaknya tetangga terdekat yang beragam yaitu $k=3$, $k=5$, dan $k=7$. Untuk mengevaluasi model yang dibuat dilakukan penghitungan performa model menggunakan modul *metrics* pada *python*.



Gambar 2. Hasil Uji Evaluasi K-NN dengan k beragam

Berdasarkan hasil evaluasi pada Gambar 2 terhadap tiap parameter k , dapat disimpulkan bahwa percobaan dengan performa terbaik diperoleh pada percobaan dengan ID Percobaan P-7 yang menggunakan skenario 90% data latih atau 9019 record data dan 10% data uji atau 1003 record data dan menggunakan nilai $k = 5$ dengan nilai akurasi sebesar 90,23%; nilai precision sebesar 90,23%, dan nilai recall sebesar 72,38%. Adapun model yang menghasilkan performa terendah diperoleh pada percobaan dengan ID P-8 dengan proporsi data latih sebesar 95% atau 9620 data ulasan dan data

Kata yang mendominasi pada sentimen netral pengguna aplikasi Maxim diantaranya kata aplikasi; driver; maxim; puas/puas; map; daftar; baik; order; drivernya; dan harga. Adapun jumlah kemunculan kata pada sentimen netral, positif, dan negatif secara lebih rinci tertera pada Tabel 12.

Tabel 12. Total Kemunculan Kata pada Ulasan Netral, Positif, dan Negatif

Netral			Positif			Negatif		
No	Word	Jumlah	No	Word	Jumlah	No	Word	Jumlah
1	aplikasi	135	1	terima kasih	1402	1	driver	275
2	driver	134	2	bagus	969	2	aplikasi	220
3	maxim	107	3	mantap	765	3	maxim	157
4	puas	100	4	baik	716	4	susah	118
5	map	62	5	cepat	509	5	order	103
6	daftar	50	6	driver	349	6	map	76
7	baik	49	7	oke	334	7	lama	75
8	order	49	8	maxim	311	8	tolong	67
9	drivernya	48	9	murah	278	9	error	63
10	harga	46	10	baik ramah	263	10	gagal	58

Berbeda halnya dengan sentimen netral, berikut merupakan visualisasi kata pada sentimen positif dari ulasan pengguna aplikasi Maxim pada Gambar 5.



Gambar 5. Wordcloud Sentimen Positif

Dapat disimpulkan bahwa 10 kata yang paling mendominasi pada sentimen positif ulasan pengguna aplikasi Maxim diantaranya kata terima kasih; bagus; mantap; baik; cepat; driver; oke; maxim; murah; dan baik ramah. Adapun visualisasi kata pada data ulasan yang bersentimen negatif tertera pada Gambar 6.



Gambar 6. Wordcloud Sentimen Negatif

Mengacu pada Gambar 5 dapat disimpulkan bahwa 10 kata yang mendominasi pada sentimen negatif yaitu kata driver; aplikasi; maxim; susah; order; map; lama; tolong; error; dan gagal. Jika dianalisis lebih lanjut, terdapat kemunculan kata yang sama antara satu kelas dengan kelas lainnya pada 10 kata dengan kemunculan tertinggi sebagaimana yang tertera pada Tabel 13.

Tabel 13. Kemunculan Kata yang Sama

No.	Kata	Sentimen		
		Netral	Positif	Negatif
1	aplikasi	√	-	√
2	driver	√	√	√
3	maxim	√	√	√
4	map	√	-	√
5	baik	√	√	-
6	order	√	-	√

Berdasarkan Tabel 13 dapat ditarik beberapa kesimpulan, yaitu:

- Terdapat 2 kata yang sering muncul dalam semua kelas sentimen yaitu kata “driver” dan “maxim”
- Kata “aplikasi”, “map”, dan “order” sering muncul pada kelas sentimen netral dan negatif
- Kata “baik” sering muncul pada kelas sentimen netral dan positif

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa performa terbaik pada percobaan dengan pembagian data latih dan data uji, serta nilai k yang beragam diperoleh pada percobaan dengan pembagian data latih 90%, data uji 10% dan menggunakan nilai $k = 5$ dengan nilai akurasi, presisi, dan recall berturut-turut sebesar 90,23%; 90,23%; dan nilai recall 72,38%. Klasifikasi sentimen dengan model terbaik menggunakan parameter $k=3$ menghasilkan 79,26% sentimen positif, 17,25% sentimen netral, dan 3,49% sentimen negatif, selain itu terdapat 2 kata dengan frekuensi kemunculan paling tinggi di ketiga kelas sentimen yaitu kata “driver” dan “maxim”.

REFERENCES

- [1] A. Popy, “RESOLUSI KONFLIK ANTARA PENGEMUDI TRANSPORTASI ONLINE DENGAN PENGEMUDI TRANSPORTASI KONVENSIONAL (Studi Hambatan Struktural Penyelesaian Konflik antara Pengemudi Gojek, Pengemudi Angkutan Kota (Angkot), Pengemudi Ojek Pangkalan di Kota Padang),” Universitas Andalas, 2018.
- [2] Google and TEMASEK, “e-Conomy SEA 2018 Southeast Asia’s internet economy hits an inflection point.” pp. 1–32, 2018.
- [3] J. P. Kompetitif, “Laporan Survei Penetrasi \& Profil Perilaku Pengguna Internet Indonesia 2018,” *J. Pemasar. Kompetitif*, vol. 1, no. 3, pp. 71–80.
- [4] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [5] M. T. Diwandanu and L. M. Wisudawati, “Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K Nearest Neighbors,” *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 1–16, 2023, doi: 10.35760/ik.2023.v28i1.7909.
- [6] 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.
- [7] B. Liu, *Sentiment analysis and opinion mining*. Springer Nature, 2022.
- [8] O. Somantri and D. Apriliani, “Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 537, 2018, doi: 10.25126/jtiik.201855867.
- [9] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [10] N. Tri Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [11] N. S. N. Salam, A. A. Supianto, and A. R. Perdanakusuma, “Analisis Sentimen Opini Mahasiswa Terhadap Saran Kuesioner Penilaian Kinerja Dosen dengan Menggunakan TF-IDF dan K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 6148–6156, 2019.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, pp. 115–123, 2020.
- [13] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [14] A. N. S. Rahayu, T. I. Hermanto, and I. M. Nugroho, “Sentiment Analysis Using K-Nearest Neighbor Based on Particle Swarm Optimization According To Sunscreen’S Reviews,” *J. Tek. Inform.*, vol. 3, no. 6, pp. 1639–1646, 2022, doi: 10.20884/1.jutif.2022.3.6.425.
- [15] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, “Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.
- [16] 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, pp. 1–13, 2018.
- [17] S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, “Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [18] R. Siringoringo and J. Jamaludin, “Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online,” *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [19] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [20] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [21] D. Era, S. Andryana, and A. Rubhasy, “Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 7, no. 1, pp. 263–272, 2023.
- [22] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, “Tweets responding to the Indonesian Government’s handling of COVID-19: Sentiment analysis using SVM with normalized poly kernel,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, p. 112, 2020.
- [23] R. Prakoso, “kbba,” 2017. <https://github.com/ramaprakoso/analisis-sentimen/blob/master/kamus/kbba.txt>.
- [24] M. ARIFAH, “Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan Lexicon-Based Dan Naive Bayes,” 2019.