

ANALISIS SENTIMEN TERHADAP TWIT MAXIM PADA TWITTER MENGGUNAKAN R PROGRAMMING DAN K NEAREST NEIGHBORS

¹Muhamad Trian Diwandanu, ²Lu'lu Mawaddah Wisudawati*

^{1,2}Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

¹triandiwandanu17@student.gunadarma.ac.id, ²lulu_mawadah@gunadarma.ac.id

*) Penulis korespondensi

Abstrak

Usaha transportasi saat ini sudah banyak yang berbasis online dalam pelayanannya seperti pemesanan, pembayaran dan pemberian ulasan. Salah satu jasa transportasi online yang sudah ada di Indonesia yaitu Maxim. Masyarakat biasanya memberikan opini mereka terhadap layanan yang diberikan oleh Maxim melalui Twitter. Twit yang ditulis oleh masyarakat pengguna Twitter merupakan sumber data yang valid untuk dilakukan analisis sentimen. Tujuan penulisan ini adalah melakukan analisis sentimen terhadap twit maxim pada Twitter. Metode klasifikasi analisis sentimen pada penulisan ini menggunakan K Nearest Neighbors (KNN) untuk mengklasifikasi data serta Lexicon-Based sebagai penentu sentimen positif, negatif dan netral. Tahapan awal pada analisis sentimen ini yaitu tahap pengambilan data, pre-processing, yang terdiri dari Filtering & Casefolding, perbaikan kata tidak baku, mengubah kata bernegasi, Stopword Removal dan penghapusan spasi berlebih. Setelah itu, dilakukan pelabelan data dan pemberian skor menggunakan metode Lexicon Based. Dataset hasil dari pre-processing dan Lexicon Based digunakan untuk proses klasifikasi menggunakan KNN. Hasil terbaik didapatkan menggunakan data latih 80% sebanyak 702 data dan data uji 20% sebanyak 175 data dengan $k=1$ dengan akurasi sebesar 95,43%.

Kata Kunci: K-Nearest Neighbors, Filtering, Case Folding, Maxim, R Programming, Analisis Sentimen

Abstract

Many transportation businesses are now based online in their services such as booking, payment and reviews. One of the online transportation services that already exist in Indonesia is Maxim. People usually give their opinions on the services provided by Maxim through Twitter. Tweets written by Twitter users are a valid source of data for sentiment analysis. The sentiment analysis classification method in this paper uses K Nearest Neighbors (KNN) to classify data and Lexicon-Based as a determinant of positive, negative and neutral sentiments. The initial stages in this sentiment analysis are the data retrieval stage, pre-processing, which consists of Filtering & Casefolding, repairing nonstandard words, changing negated words, Stopword Removal and removing excess spaces. After that, data labeling and scoring is done using the Lexicon Based method. Datasets resulting from pre-processing and Lexicon Based are used for the classification process using KNN. The best results were obtained using 80% training data of 702 data and 20% test data of 175 data with $k = 1$ with an accuracy of 95.43%.

Keywords: K-Nearest Neighbors, Filtering, Case Folding, Maxim, R Programming, Sentiment Analysis

PENDAHULUAN

Seiring dengan perkembangan teknologi, usaha dibidang transportasi cukup berkembang di Indonesia. Usaha transportasi saat ini sudah berbasis online dalam pelayanannya seperti pemesanan, pembayaran dan pemberian ulasan. Hal ini mempermudah masyarakat menggunakan jasa transportasi dalam aktivitasnya sehari-hari. Aplikasi transportasi online ini berbasis aplikasi yang menggunakan media internet sebagai penghubung dan penyedia layanan. Semakin banyaknya jasa transportasi online ini berdampak pada semakin ketatnya persaingan. Faktor yang menjadi penyebab terjadinya hal tersebut yaitu adanya persaingan harga dan kualitas pelayanan. Salah satu jasa transportasi online yang sudah ada di Indonesia yaitu Maxim.

Maxim berasal dari perusahaan Rusia yang bernaung dalam PT. Teknologi Perdana Indonesia yang hadir di Indonesia sejak Juli 2018. Masyarakat biasanya memberikan opini mereka terhadap layanan yang diberikan oleh Maxim melalui Twitter. Twitter merupakan salah satu aplikasi sosial media yang masih banyak digunakan saat ini untuk membagikan apa yang terjadi dalam kehidupan sehari-hari penggunaanya. Twit yang ditulis oleh masyarakat pengguna Twitter merupakan sumber data yang valid untuk dilakukan analisis sentimen. Analisis sentimen merupakan suatu teknik pengestrakan data dari sebuah teks untuk mengambil informasi

mengenai sentimen yang bernilai positif, netral, maupun negatif [5].

Banyak penelitian yang menggunakan sentiment analysis dalam berbagai aplikasi, salah satunya yaitu aplikasi twitter menggunakan metode Machine Learning [1]. Metode tersebut juga digunakan untuk menganalisis COVID-19 dengan metode advance machine learning [2]. Metode lain yang digunakan dalam sentiment analisis yaitu Algoritma Naive Bayes dan Support Vector Machine dalam menganalisis tokoh publik [14] dan menganalisis sentiment transportasi online [4]. Penelitian mengenai analisis sentimen salah satunya dilakukan oleh Ardianda Aryo Prakoso dkk [2] pada amazon web review dengan mengambil dataset sebanyak 1000 data menggunakan UCI Machine Learning, setelah melalui tahap *pre-processing* melakukan metode Lexicon dimana mengubah kata imbuhan menjadi kata dasar kemudian diekstraksi menjadi kata sentimen yang berbobot, dan yang terakhir menghitung akurasi dari tiga metode klasifikasi antara lain Decision Tree, K-Nearest Neighbor dan Support Vector Machine (SVM). Hasil akurasi terbaik yang didapatkan menggunakan metode Lexicon adalah metode K-Nearest Neighbor dengan nilai akurasi 92,67% kemudian metode SVM dengan akurasi 91,33% dan Decision Tree dengan akurasi 82%.

Penelitian selanjutnya dilakukan oleh Risma Putri Fitrianti dkk [12] dengan melakukan implementasi algoritma K- Nearest

Neighbor terhadap analisis sentimen review restoran dengan teks Bahasa Indonesia. Penelitian tersebut menggunakan data latih sebesar 70% dari dataset yaitu sebanyak 140 data dan data uji sebesar 30% dari dataset sebanyak 59 data. Peneliti melakukan pengujian menggunakan confusion matrix dan didapatkan hasil akurasi dari analisis sentimen terhadap review restoran dengan teks Bahasa Indonesia menggunakan K - Nearest Neighbor dengan persentase 96,61%, nilai $k = 1$ dan nilai error rate sebesar 3,39%. Penelitian selanjutnya dilakukan oleh Mst. Tuhin Akter [9] dkk dengan melakukan perbandingan antara algoritma Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) dan XGBoost pada review produk *E-commerce*. Hasil penelitian menunjukkan bahwa KNN memiliki akurasi yang paling tinggi diantara lainnya. Algoritma KNN memiliki akurasi sebesar 96,25%, SVM memiliki akurasi sebesar 94,35%, Random Forest memiliki akurasi sebesar 90,84%, XGBoost memiliki akurasi sebesar 90,56% dan Logistic Regression memiliki akurasi sebesar 90,33%.

Rani Puspita dan Agus Widodo [11] melakukan perbandingan Metode KNN, Decision Tree, dan Naïve Bayes terhadap analisis sentimen pengguna layanan BPJS. Dalam penelitiannya menggunakan RapidMiner untuk pengambilan data Twitter sebanyak 1000 data. Hasil yang didapatkan yaitu akurasi dengan metode KNN sebesar 95,58%, Decision Tree memiliki akurasi

sebesar 96,13% dan Naive Bayes memiliki 89,14%. Peneliti selanjutnya Mhd. Furqan dkk [8] melakukan analisis sentimen menggunakan K-Nearest Neighbor terhadap new normal masa Covid-19 di Indonesia. Peneliti menggunakan data latih sebesar 80% dan data uji 20% dari total dataset sebanyak 1000 data yang diambil dari Twitter. Hasil klasifikasi K-NN dengan nilai $k = 1$ menghasilkan pengujian use training set dengan accuracy sebesar 100%, 92,60% untuk 10-fold cross-validation dan 94,50% untuk 80% percentage split.

Berdasarkan penelitian-penelitian sebelumnya [2,12,9,12,11,8], peneliti mengembangkan penelitian dengan perbaikan klasifikasi sentimen yang lebih menunjukkan sentimen tersebut ke arah positif atau negatif, sehingga sentimen yang bernilai netral dapat diminimalisir, kemudian menggunakan metode klasifikasi sentimen seperti Lexicon Based agar pemberian kategori sentiment positif, netral negatif dilakukan secara otomatis dengan jumlah data yang diolah dan menambahkan model data latih lebih banyak sehingga meningkatkan tingkat akurasi dari analisis yang dilakukan. Selain itu, peneliti mengembangkan metode klasifikasi sentimen untuk labeling data secara otomatis dan juga melakukan penambahan kata pada kamus Lexicon dan kamus bahasa *slang*. Pemilihan objek penelitian twit Maxim dikarenakan saat ini masyarakat pengguna Twitter banyak yang mengulas serta bertukar opini mengenai layanan aplikasi transportasi online Maxim. Banyaknya twit memungkinkan dilakukan

analisa untuk dilakukan penelitian. Berdasarkan latar belakang tersebut, tujuan dari penelitian ini yaitu melakukan analisis sentiment terhadap twit maxim menggunakan R programming dan algoritma K-NN dengan harapan dapat membantu perusahaan maxim dalam meningkatkan layanan berdasarkan review pelanggan.

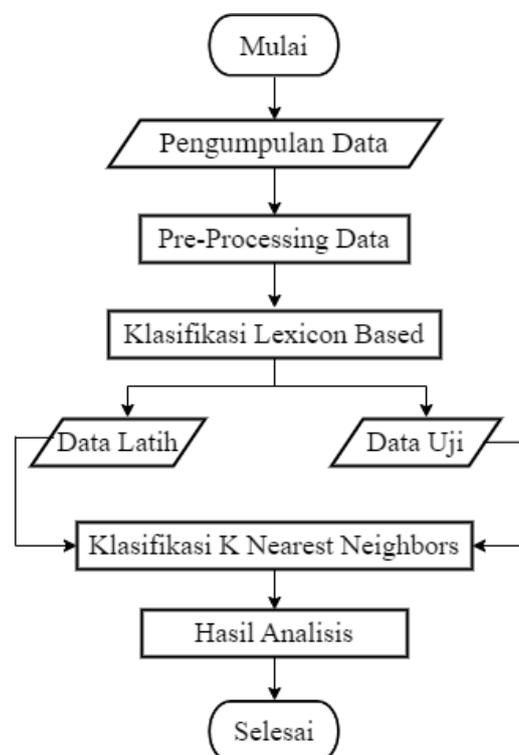
METODE PENELITIAN

GAMBARAN UMUM SISTEM

Penelitian ini dilakukan dengan beberapa tahapan yaitu pengumpulan data, *pre-processing* data, klasifikasi *Lexicon Based*, pembagian data latih dan data uji, klasifikasi *K*

Nearest Neighbors dan hasil analisis (lihat Gambar 1). Gambar 1. merupakan flowchart dari gambaran umum sistem.

Pengumpulan data dilakukan dengan menggunakan RStudio dengan API Twitter, kemudian tahapan *pre-processing* merupakan proses dimana data diolah sebelum dilakukan analisis. Tahapan klasifikasi *Lexicon Based* dilakukan dengan mengubah kata imbuhan menjadi kata dasar kemudian diekstraksi menjadi kata sentimen yang berbobot dan hasil klasifikasi dibagi menjadi menjadi data latih dan data uji, setelah itu dilakukan klasifikasi menggunakan *K Nearest Neighbors* untuk menghitung akurasi dari hasil analisis.



Gambar 1. Gambaran Umum Sistem

Pengumpulan Data

Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming dan K Nearest Neighbors pada langkah melakukan pengumpulan data twit maxim melalui media sosial Twitter menggunakan RStudio dengan API Twitter. Data yang digunakan adalah twit pengguna Twitter yang berbahasa Indonesia dan diperoleh secara random sebanyak 1639 data. Berikut adalah data yang telah diperoleh dengan menggunakan kata kunci “maxim” dapat dilihat pada Tabel 1.

Pre-processing

Tahap *pre-processing* merupakan tahap yang paling penting untuk membuat analisis sentimen. Ditahap ini merubah data mentah menjadi data yang bersih untuk diolah dan

berbobot agar dapat diproses menjadi hasil yang terbaik dengan menghapus data atau kalimat yang sama. Jika terdapat kalimat yang sama akan dihapus pada proses *pre-processing* ini agar tidak ada redundansi data. Tahapan dari proses *pre-processing* dapat dilihat pada Gambar 2. Gambar 2 merupakan tahapan *pre-processing* yang terdiri dari penghapusan dan penormalan pada data Twit dari elemen yang tidak akan digunakan dalam proses analisis sentiment seperti link, tagar, simbol dan elemen lainnya yang biasanya ada pada Twit. Kemudian mengganti kata tidak baku, memperbaiki kata negasi dan menghilangkan spasi berlebih pada data. Tahapan pada langkah *pre-processing*, yaitu : *Filtering*, *Case Fold*, Perbaikan kata tidak baku, mengubah kata bernegasi, *stopwrod removal*, dan penghapusan spasi berlebih.

Tabel 1. Contoh Data Twit

No	Text
1	Puji Tuhan, alhamdulillah, amitohu, rahayu. Hari ini belum dipercayakan dapat orderan dari pagi. Sementara Maxim da... https://t.co/zgG0tZl9ij
2	@convomfs Tp maxim tuh klo nunggu suka lama bgt anjing bisa ampe 1 jam bgst
3	Alhamdulillah, sejauh ini gue g pernah dpat driver maxim yg modelan gini
4	Maxim lamaaa bgt anjir
5	Maxim nih udh tw g ada brngny tpi tokony g ada tulisan tutup. Trus g bisa cancel dong anjng susah bngt
6	Maxim bad bngt ewhhh, mana abangny galak bbgt
7	@prozmosi Kenapa temen-temen gw pada pake maxim ya
8	Maxim emang gitu. Bukannya liat maps malah marah-marah ke gue karena nyasar, padahal gue juga baru pertama ke daera... https://t.co/sIGU2xgPiL
9	@convomfs dua kali nyoba maxim pengalamannya gaenak semua
10	@convomfs Maxim murah tapi kualitas drivernya nol banget



Gambar 2. Tahapan *Pre-processing*

Algoritma 1. Filtering

Input: Dataset dalam bentuk twit

Output: Hasil teks tidak terdapat symbol, tanda baca dan angka

1. *Mulai*
 2. *Baca input*
 3. *Jika Teks terdapat symbol, tanda baca dan angka Maka Teks dihilangkan. Jika tidak, Teks disimpan.*
 4. *Selesai*
-

Algoritma 2. Case Fold

Input: Teks hasil filtering

Output: Teks hasil case fold

1. *Mulai*
 2. *Baca input*
 3. *Mengubah Huruf Kapital menjadi Huruf Kecil*
 4. *Simpan Teks*
 5. *Selesai*
-

Filtering

Filtering merupakan tahap untuk menghilangkan kata-kata yang tidak akan dipakai untuk analisis sentimen pada dataset yang ada seperti simbol-simbol, tanda baca dan juga angka. Tahapan filtering dapat dilihat pada Algoritma 1. Contoh teks sebelum filtering adalah “lagi nunggu maxim, notif

muslim pro muncul <https://t.co/nJQMH0c7qL>”. Hasil Teks setelah dilakukan tahap *filtering* adalah “lagi nunggu maxim notif muslim pro muncul”.

Case fold

Tahapan kedua dari *pre-processing* adalah proses *case fold* (lihat Algoritma 2).

Algoritma 2 menjelaskan terkait tahapan *case fold* yaitu dengan mengambil dataset hasil dari *filtering* yang akan dipakai, kemudian mengubah semua huruf kapital menjadi huruf kecil dan hasilnya adalah data yang sudah tidak ada huruf kapital.

Perbaikan Kata Tidak Baku

Tahapan ketiga dari *pre-processing* adalah proses perbaikan kata tidak baku. Tahapan dari proses perbaikan kata tidak baku dapat dilihat pada Algoritma 3. Algoritma 3 menjelaskan Langkah yang dilakukan dalam proses merubah kata tidak baku yaitu dengan mencocokkan kata tidak baku dengan kata baku yang telah di konversi pada lexicon.

Langkah awal perbaikan kata tidak baku yaitu merubah kata tidak baku menjadi kata baku. Kata yang tidak baku seperti singkatan atau istilah seperti (“Slalu”, “ajah”) dan juga

bisa menjadi bahasa gaul seperti (“loe”, “letoy”). Perubahan bahasa gaul menggunakan bahasa harian, menggunakan Bahasa Indonesia.

Bahasa harian Indonesia banyak terdapat kata-kata tidak baku yang dipasangkan dengan kata bakunya, sampel dari lexicon dapat dilihat pada Tabel 2.

Perubahan Kata Negasi

Tahapan keempat dari *pre-processing* adalah proses perbaikan kata tidak baku. Pada tahap ini mengubah kata bernegasi yaitu mengubah kata-kata yang bermakna sangkalan seperti “tidak” dan “ada” ditambah dengan tanda *underscore* (_) diantara kata negasi tersebut agar mendapat makna yang sebenarnya Tahap perubahan kata negasi dapat dilihat pada Gambar 3.

Algoritma 3. Perbaikan kata tidak baku

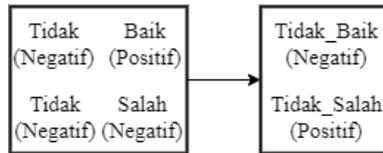
Input: Teks hasil *case fold*

Output: Teks hasil perbaikan kata tidak baku

1. Mulai
 2. Baca input
 3. Pengecekan kata tidak baku dalam database slang Lexicon
Jika terdapat kata/teks tidak baku, maka kata/teks diubah menjadi baku.
Jika tidak terdapat kata/teks tidak baku, maka kata/teks disimpan
 4. Simpan Teks
 5. Selesai
-

Tabel 2. Contoh Lexicon

Kata Slang	Kata Formal
ak	aku
bgt	banget
Tdk	tidak
Karna	karena
ttp	tetap



Gambar 3. Perubahan Kata Negasi

Algoritma 4. Stop removal

Input: Teks hasil perubahan kata negasi

Output: Teks hasil stop removal

1. *Mulai*
 2. *Baca input*
 3. *Pengecekan stopword dalam database kamus stopword*
Jika terdapat stopword, maka dihapus
Jika tidak terdapat stopword, maka disimpan
 4. *Simpan Teks*
 5. *Selesai*
-

Tabel 3. Contoh Kamus Stopword

<u>No</u>	<u>Stopword</u>
1	Ada
2	Adalah
3	Bahkan
4	Bahwa
5	Dan

Stopword Removal

Tahapan kelima *pre-processing* adalah proses *stopword removal* (lihat Algoritma 4). Algoritma 4 merupakan tahapan stop removal, dimana pada langkah *stopword removal* dilakukan penghapusan kata sambung, kata depan dan juga kata ganti. Proses ini dilakukan dengan mencari kata pada data dalam kamus

stopword (lihat Table 3), jika kata tersebut ditemukan pada kamus maka akan dihilangkan.

Penghapusan Spasi Berlebih

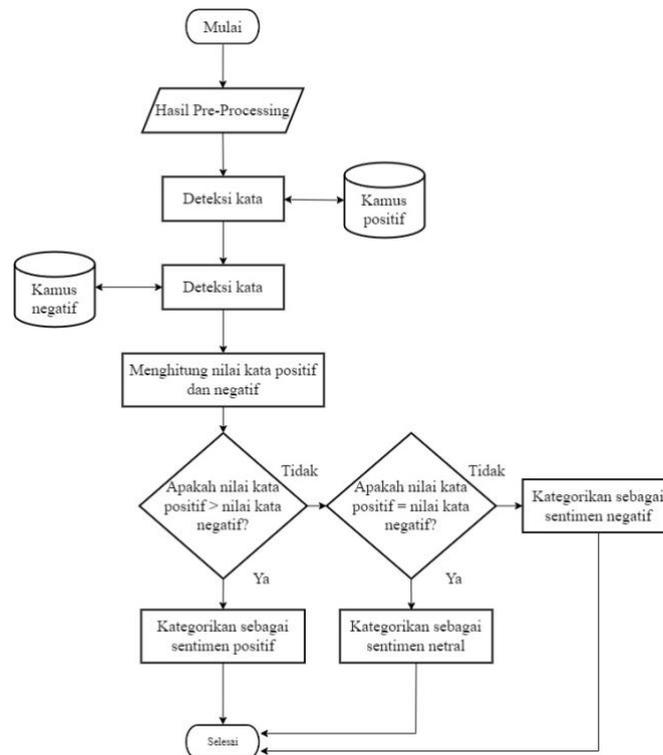
Pada tahap penghapusan spasi berlebih, data yang terdapat lebih dari satu spasi akan dihapus sehingga hanya akan menjadi satu spasi saja. Tahapan dari proses penghapusan spasi berlebih dapat dilihat pada Algoritma 5.

Algoritma 5. Penghapusan Spasi Berlebih

Input: Teks hasil stop removal

Output: Teks hasil penghapusan spasi berlebih

1. *Mulai*
 2. *Baca input*
 3. *Pengecekan spasi berlebih pada teks*
Jika terdapat spasi berlebih, maka dihapus
Jika tidak terdapat spasi berlebih, maka disimpan
 4. *Simpan Teks*
 5. *Selesai*
-



Gambar 4. Tahap Klasifikasi Lexicon Based

Rancangan Klasifikasi Sentimen *Lexicon Based*

Metode *Lexicon Based* akan membandingkan masing-masing data uji pada data kata sentimen Bahasa Indonesia dan akan menghasilkan sebuah nilai sentimen. Tahapan dari proses klasifikasi *Lexicon Based* dapat dilihat pada Gambar 4. Gambar 4 merupakan tahapan klasifikasi *Lexicon Based*.

Klasifikasi ini akan mencocokkan kata-kata yang ada pada kamus positif dan kamus negatif. Jika data twit dalam satu kalimat terdapat jumlah kata sentimen positif yang lebih besar dibandingkan negatif maka akan dikategorikan sebagai sentimen positif, sedangkan kata sentimen negatif dalam satu kalimat lebih besar dibandingkan positif, maka kalimat tersebut dikategorikan sebagai sentimen negatif

sedangkan kata positif dan negatif dalam satu kalimat memiliki nilai sama, maka dikategorikan sebagai netral. Klasifikasi *K Nearest Neighbors* menggunakan persamaan berikut:

$$Cos(\theta_{QD}) = \frac{\sum_{i=1}^n Q_i D_i}{\sqrt{\sum_{i=1}^n (Q_i)^2} \cdot \sqrt{\sum_{i=1}^n (D_i)^2}} \quad (1)$$

Persamaan (1) $Cos(\theta_{QD})$ merupakan kemiripan *Q* terhadap dokumen *D*, variable *Q* merupakan data uji, variable *D* merupakan data latih dan variable *n* merupakan banyaknya data.

UJI AKURASI

Confusion matrix digunakan untuk melakukan uji coba pengukuran akurasi. Tujuan dari pengukuran ini adalah untuk mengetahui tingkat akurasi dari klasifikasi data twit yang dijalankan oleh sistem menggunakan metode klasifikasi *Lexicon*

Based dan K Nearest Neighbors. Untuk menghitung akurasi dari Confusion Matrix dapat menggunakan persamaan (2) berikut:

Accuracy =

$$\frac{TN (True Negative)+TNe (True Netral)+TP (True Positive)}{TN+FN (False Negative)+TNe+FNe (False Netral)+TP+FP (False Positve)}$$

(2)

HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dari twit pengguna Twitter yang berbahasa Indonesia dan diperoleh secara random sebanyak 1639 data dan hasil klasifikasi menggunakan metode Lexicon Based yang telah dilakukan sebelumnya yaitu menjadi 1000 data untuk nantinya dijadikan sebagai data latih atau data

uji. Kemudian kolom 10 entiment akan digunakan untuk mendapatkan hasil mengenai sentiment positif dan 10 entimen. Pengujian dilakukan dengan 3 skema yang dapat dilihat

pada Tabel 4.

Hasil Pre-processing

Tahap *filtering* merupakan tahap untuk menghilangkan karakter yang tidak digunakan untuk 10 entiment analisis seperti 10 entim-simbol, tanda baca dan juga angka. Hasil tahap *filtering* analisis 10 entiment terhadap twit maxim dapat dilihat pada Tabel 5.

Pada tahap *casefold*, semua data yang memiliki huruf kapital akan diubah menjadi huruf kecil. Hasil tahap *casefold* analisis sentimen terhadap twit maxim dapat dilihat pada Tabel 6.

Tabel 4. Skema Pengujian

Skema	Data latih	Data Uji
1	80%	20%
2	75%	25%
3	70%	30%

Tabel 5. Hasil Filtering

No	Input	Output
1	lagi nunggu maxim, notif muslim pro muncul https://t.co/nJQMh0c7qL	lagi nunggu maxim notif muslim pro muncul
2	pls ini driver maxim w udah nunggu setengah jem ga dateng2	pls ini driver maxim w udah nunggu setengah jem ga dateng
3	maxim food tuh emang lama banget kah? Harus konfir lewat wa telfon dan chat?	Maxim food tuh emang lama banget kah harus konfir lewat wa telfon dan chat

Tabel 6. Hasil Case Fold

No	Input	Output
1	Maxim lagi error	maxim lagi error
2	Aku baru tau klo maxim itu aplikasi ojol dri russia	aku baru tau klo maxim itu aplikasi ojol dri russia
3	Gawat kali nunggu maxim setengah jam	gawat kali nunggu maxim setengah jam

Tahap perbaikan kata tidak baku merupakan tahap yang mengubah kata tidak baku menjadi kata baku seperti singkatan atau bahasa gaul. Hasil tahap perbaikan kata tidak baku analisis sentimen terhadap twit maxim dapat dilihat pada Tabel 7. Pada tahap mengubah kata bernegasi yaitu mengubah kata-kata yang bermakna sangkalan seperti “tidak” dan “ada” ditambah dengan tanda (_) diantara kata negasi tersebut agar mendapat makna yang sebenarnya. Hasil tahap mengubah kata bernegasi analisis sentimen terhadap twit maxim dapat dilihat pada Tabel 8. Tahap

stopword removal merupakan tahap yang menghapus kata-kata yang tidak digunakan seperti kata sambung, kata depan dan juga kata ganti sesuai dengan kamus *stopword* yang digunakan. Hasil tahap *stopword removal* analisis sentimen terhadap twit maxim dapat dilihat pada Tabel 9. Pada tahap penghapusan spasi berlebih, data yang terdapat lebih dari satu spasi akan dihapus sehingga hanya akan menjadi satu spasi saja. Hasil tahap penghapusan spasi berlebih analisis sentimen terhadap twit maxim dapat dilihat pada Tabel 10.

Tabel 7. Hasil Perbaikan Kata Tidak Baku

No	Input	Output
1	gw naik maxim td harganya murah bgt cuma k mungkin karna pengguna baru nah trs orang nya bilang ga mungkin	gue naik maxim tadi harganya murah banget cuma kak mungkin karena pengguna baru nah terus orang nya bilang tidak mungkin
2	asli ini mah ga nyaman bgt malag mending ttp pake aplikasi aja kalo ada apa bisa lgsg lapor	asli ini mah tidak nyaman banget malag mending tetap pakai aplikasi saja kalo ada apa bisa langsung lapor
3	ko gw gak bisa daftar maxim ya susah bngt	kok gue tidak bisa daftar maxim ya susah banget

Tabel 8. Hasil Mengubah Kata Bernegasi

No	Input	Output
1	gue naik maxim tadi harganya murah banget cuma kak mungkin karena pengguna baru nah terus orang nya bilang tidak mungkin	gue naik maxim tadi harganya murah banget cuma kak mungkin karena pengguna baru nah terus orang nya bilang tidak_mungkin
2	asli ini mah tidak nyaman banget malag mending tetap pakai aplikasi saja kalo ada apa bisa langsung lapor	asli ini mah tidak_nyaman banget malag mending tetap pakai aplikasi saja kalo ada apa bisa langsung lapor
3	kok gue tidak bisa daftar maxim ya susah banget	kok gue tidak_bisa daftar maxim ya susah banget

Tabel 9. Hasil Stop Removal

No	Input	Output
1	benar banget pernah memakai maxim di jogja terus supirnya mengambek anjir terus kembalian gue gadikasih	banget memakai maxim jogja supirnya mengambek anjir kembalian gue gadikasih
2	maxim memang begitu bukannya lihat maps malah marahmarah ke gue karena nyasar padahal gue juga baru pertama ke	maxim lihat maps marahmarah gue nyasar gue
3	maxim memang tidak_bisa bayar cashless bang cuma bisa cash jadi siapin saja duit pas	maxim tidak_bisa bayar cashless bang cash siapin duit pas

Tabel 10. Hasil Penghapusan Spasi Berlebih

No	Input	Output
1	sumpah tidak_mengerti maxim drivernya sinting banget gue teror	Sumpah tidak_mengerti maxim drivernya sinting banget gue teror
2	ih gue chat driver maxim gue block sih ya risih berkali kali	ih gue chat driver maxim gue block sih ya risih berkali kali
3	tidak_kalau indonesia bahasa nyaman bicara bahasa	tidak_kalau indonesia bahasa nyaman bicara bahasa

Tabel 11. Hasil Klasifikasi *Lexicon Based*

Twit	Positif	Negatif	Hasil	Sentimen
nama maxim guys penipu kena tipu pesanan nipu perwakilan	0	2	-2	Negatif
btw mas mas maxim kadang seram anjrit	0	1	-1	Negatif
alhamdulillah gue tidak_pernah dpat driver maxim modelan	1	0	1	Positif
maxim seram anjir drivernya	0	1	-1	Negatif

Hasil	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Akurasi	0.9543	0.7314	0.7257	0.68	0.68	0.6742	0.68	0.6685	0.6685	0.6857
Sensitivitas (Positif)	0.9286	0.4642	0.3214	0.2857	0.2142	0.2142	0.2142	0.1785	0.1785	0.1785
Sensitivitas (Netral)	0.9823	0.8938	0.9292	0.9381	0.9381	0.9381	0.9469	0.9469	0.9646	0.9735
Sensitivitas (Negatif)	0.8824	0.4118	0.3823	0.1470	0.2058	0.1764	0.1764	0.1470	0.0882	0.1470
Spesifisitas (Positif)	0.9864	0.93878	0.9659	0.9659	0.9659	0.9659	0.9727	0.9727	0.9795	0.9863
Spesifisitas (Netral)	0.9194	0.4839	0.3710	0.2097	0.2097	0.1935	0.1935	0.1613	0.1452	0.1613
Spesifisitas (Negatif)	0.9929	0.9574	0.9716	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9929

Gambar 5. Hasil Pengujian Skema 1 (Akurasi, Sensitivitas dan Spesifisitas)

Hasil Klasifikasi Sentimen *Lexicon Based*
 Metode klasifikasi *Lexicon Based* ini akan mencocokkan kata-kata yang ada pada kamus positif dan kamus negatif. Hasil dari klasifikasi menggunakan metode *Lexicon Based* dapat dilihat pada Tabel 11.

Hasil Uji Akurasi Analisis

Proses untuk pengujian skema satu menggunakan 175 data uji. Data uji tersebut dibandingkan menggunakan hasil klasifikasi yang dilakukan oleh sistem. Tabel hasil perbandingan penggunaan jumlah k pada klasifikasi KNN dengan menggunakan data latih 80% dan data uji 20% dapat dilihat pada Gambar 5. Proses untuk pengujian skema dua menggunakan 219 data uji. Data uji tersebut dibandingkan menggunakan hasil klasifikasi

yang dilakukan oleh sistem. Tabel hasil perbandingan penggunaan jumlah k pada klasifikasi KNN dengan menggunakan data latih 75% dan data uji 25% dapat dilihat pada Gambar 6. Proses untuk pengujian skema tiga menggunakan 263 data uji.

Data uji tersebut dibandingkan menggunakan hasil klasifikasi yang dilakukan oleh sistem. Tabel hasil perbandingan penggunaan jumlah k pada klasifikasi KNN dengan menggunakan data latih 70% dan data uji 30% dapat dilihat pada Gambar 7. Dari ketiga skema yang telah dilakukan (Gambar 5, Gambar 6, Gambar 7), akurasi yang tertinggi yaitu pada saat penggunaan data latih sebesar 80% dan data uji sebesar 20% dengan menggunakan k sebesar 1 dengan akurasi sebesar 0.9543. Hasil akurasi terbesar

didapatkan melalui tabel confusion matrix. Tabel confusion matrix dapat dilihat pada Tabel 11.

Dari hasil perhitungan akurasi manual diatas menggunakan persamaan (2), didapatkan hasil akurasi sebesar 95,42857% atau 95,43%

Hasil Perbandingan Penelitian Terdahulu

Hasil perbandingan penelitian terdahulu yaitu pada peneliti [2] dan [8] menggunakan metode klasifikasi sentimen lexicon based dan metode klasifikasi KNN mendapatkan nilai akurasi masing-masing sebesar 92.7% dan 94.5%. Peneliti [9], [11] dan [12] menggunakan metode klasifikasi sentimen secara manual dan

metode klasifikasi KNN mendapatkan nilai akurasi masing-masing sebesar 96.25%, 96.01% dan 96.61%. Penelitian ini menggunakan metode *pre-processing* yaitu *filtering* dan *casefold* dengan objek penelitian twit pengguna layanan maxim pada twitter. Selain itu pada penelitian ini menggunakan metode klasifikasi sentimen Lexicon Based dan metode klasifikasi KNN dengan jumlah data 1639. Hasil akurasi yang didapatkan sebesar 95,43%. Dari hasil akurasi tersebut terdapat peningkatan nilai akurasi dari peneliti-peneliti sebelumnya. Hal tersebut menunjukkan metode yang digunakan mampu memberikan hasil yang optimal.

Hasil	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Akurasi	0.9223	0.7260	0.7077	0.6940	0.6894	0.6803	0.6666	0.6803	0.6712	0.6849
Sensitivitas (Positif)	0.8421	0.5263	0.3684	0.3421	0.3421	0.3421	0.3157	0.2894	0.2631	0.2894
Sensitivitas (Netral)	0.9635	0.8832	0.9197	0.9343	0.9270	0.9197	0.9270	0.9489	0.9489	0.9562
Sensitivitas (Negatif)	0.8636	0.4090	0.3409	0.2500	0.2500	0.2272	0.1590	0.1818	0.1590	0.1818
Spesifisitas (Positif)	0.9834	0.9392	0.9558	0.9558	0.9723	0.9668	0.9668	0.9834	0.9834	0.9889
Spesifisitas (Netral)	0.8780	0.5122	0.3902	0.3049	0.2927	0.2805	0.2317	0.2317	0.2073	0.2317
Spesifisitas (Negatif)	0.9771	0.9485	0.9657	0.9885	0.9714	0.9714	0.9771	0.9771	0.9771	0.9771

Gambar 6. Hasil Pengujian Skema 2 (Akurasi, Sensitivitas dan Spesifisitas)

Hasil	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Akurasi	0.8631	0.7110	0.6692	0.6653	0.6463	0.6501	0.6425	0.6463	0.6577	0.6539
Sensitivitas (Positif)	0.8667	0.5111	0.4444	0.4000	0.4000	0.3777	0.3777	0.3333	0.3111	0.3111
Sensitivitas (Netral)	0.9299	0.9108	0.9236	0.9363	0.9363	0.9363	0.9299	0.9427	0.9618	0.9618
Sensitivitas (Negatif)	0.6885	0.3442	0.1803	0.1639	0.1147	0.1147	0.0983	0.1147	0.1311	0.1147
Spesifisitas (Positif)	0.9587	0.9541	0.9541	0.9633	0.9678	0.9678	0.9724	0.9770	0.9816	0.9862
Spesifisitas (Netral)	0.7830	0.4717	0.3208	0.2736	0.2358	0.2264	0.2170	0.2075	0.2075	0.1981
Spesifisitas (Negatif)	0.9802	0.9505	0.9752	0.9851	0.9752	0.9851	0.9752	0.9802	0.9901	0.9851

Gambar 7. Hasil Pengujian Skema 3 (Akurasi, Sensitivitas dan Spesifisitas)

Tabel 11. Confusion Matrix

		Sebenarnya		
		Prediksi Negative	Neutral	Positive
Prediksi	Negative	30	1	0
	Neutral	3	111	2
	Positive	1	1	26

KESIMPULAN DAN SARAN

Pada penelitian ini analisis sentimen telah dilakukan pada twit berbahasa Indonesia dengan kata kunci “maxim”.

Dataset yang digunakan berasal dari Twitter dengan bahasa Indonesia dan menggunakan kata kunci “maxim”. Tahapan *pre-processing*, yang terdiri dari *Filtering & Casefolding*, perbaikan kata tidak baku, mengubah kata bernegasi, Stopword Removal dan penghapusan spasi berlebih. Tahapan *Lexicon Based* untuk memberikan nilai dan kategori sentimen seperti positif, netral dan negatif. Tahapan klasifikasi menggunakan algoritma KNN. Hasil pengujian dilakukan dengan 3 skema, skema satu menggunakan 80% data latih dan 20% data uji.

Skema dua menggunakan 75% data latih dan 25% data uji. Skema 3 menggunakan 70% data latih dan 30% data uji. Ketiga skema menggunakan nilai k yang berbeda yaitu k1 sampai dengan k10. Hasil akurasi terbaik didapatkan saat menggunakan data latih 80% sebanyak 702 data dan data uji 20% sebanyak 175 data dengan k=1 yaitu 95,43%.

Pada penelitian ini terdapat beberapa hal yang dapat diperbaiki seperti menggunakan dataset yang lebih baik dan kamus *Lexicon Based* yang terbaru karena dapat mempengaruhi hasil akhir dari akurasi analisis. Algoritma KNN memiliki kelebihan yaitu efektif saat menggunakan data latih yang cukup banyak dibandingkan dengan metode lain, selain itu dataset pada penelitian ini

terdapat *noise* (data yang menyimpang dari yang diharapkan) yang cukup banyak sehingga penggunaan algoritma KNN sangat tepat untuk digunakan.

Pada metode analisis lainnya seperti Naive Baiyes dan SVM dapat digunakan untuk membandingkan metode mana yang memiliki akurasi lebih tinggi dengan menggunakan dataset yang sama.

DAFTAR PUSTAKA

- [1]. A. Ikram, M. Kumar and G. Munjal, "Twitter Sentiment Analysis using Machine Learning", 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 629-634, 2022. doi: 10.1109/Confluence52989.2022.9734154, 2022
- [2]. A. Aryo Prakoso, B. Winantesa Yananta, A. Fitra Setyawan and Muljono, "A Lexicon-Based Sentiment Analysis for Amazon Web Review," International Seminar on Application for Technology of Information and Communication, Semarang, Indonesia, pp. 503-508, doi: 10.1109/ISEMANTIC.2018.8549812, 2018.
- [3]. A. P. Giovani, Ardiansyah, Ardiansyah, T. Haryanti, L. Haryanti, & W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi", *Jurnal Teknoinfo*, Vol. 14, Pages 115, No. 2, 2020. doi: 10.33365/jti.v14i2.679, 2020.

- [4]. E. dwianto, & M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine", *Jurnal Ilmiah Teknik Informatika*, Vol. 10, Pages. 94, 2021. doi: 10.22441/format.2021.v10.i1.009, 2021.
- [5]. F. V. Sari & A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi", *Jurnal SIMETRIS*, Vol 10, No. 2, pp 681-686, 2019.
- [6]. H. Sudira, A.L. Diar, & Y. Ruldeviyani, "Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia", *International Workshop on Big Data and Information Security (IW BIS)*, pp. 21-26, 2019. doi: 10.1109/IWBIS.2019.8935700, 2019.
- [7]. M. S. Mustaqbal, R. F. Firdaus, and H. Rahmadi, "Pengujian Aplikasi Menggunakan Black Box Testing Boundary Value Analysis (Studi Kasus : Aplikasi Prediksi Kelulusan SNMPTN)", *Jurnal Ilmiah Teknologi Informasi Terapan* Vol 1, No 3, pp 31-36, 2015. doi:10.33197/jitter.vol1.iss3.2015.62, 2015
- [8]. M. Furqan, Sriani, & S. M. Sari. "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia", *Jurnal Teknologi Informasi (Techno.COM)*, Vol 21, No.1, pp 52-61, 2022. doi: <https://doi.org/10.33633/tc.v21i1.5446>, 2022.
- [9]. M. T. Akter, M. Begum and R. Mustafa, "Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors," *International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, Dhaka, Bangladesh, pp. 40-44, doi: 10.1109/ICICT4SD50815.2021.9396910, 2021.
- [10]. N. Yadav, O. Kudale, S. Gupta, A. Rao and A. Shitole, "Twitter Sentiment Analysis Using Machine Learning For Product Evaluation," *International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, pp. 181-185, doi: 10.1109/ICICT48043.2020.9112381, 2020.
- [11]. R. Puspita dan A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS", *Jurnal Informatika Universitas Pamulang*, Vol 5, No. 4, pp 646-654, 2020. doi: <http://dx.doi.org/10.32493/informatika.v5i4.7622>, 2020.

- [12]. R. P. Fitrianti, A. Kurniawati, and D. Agustien, "Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan *Teks Bahasa Indonesia," Seminar Nasional Aplikasi Teknologi Informasi , pp. 27–32, 2019.
- [13]. S. Andhale, P. Mane, M. Vaingankar, D. Karia and K. T. Talele, "Twitter Sentiment Analysis for COVID-19," International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2021, pp. 1-12, doi: 10.1109/ICCICT50803.2021.9509933. 2021.
- [14]. Widowati, Tanthy & Sadikin, Mujiono, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine, Simetris", *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol 11, pp: 626-636, doi: 10.24176/simet.v11i2.4568, 2021.
- [15]. Y. Heriyanto, "Perancangan Sistem Informasi Rental Mobil Berbasis Web Pada Pt.Apm Rent Car", *Jurnal Intra Tech*, Vol 2, No. 2, 2018.
- [16]. Y.N. Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *Jurnal Fasilkom*, vol. 11, no. 2, pp. 97–103, 2021.