

ANALISIS SENTIMEN DAN KLASIFIKASI TWEETS BERBAHASA INDONESIA TERHADAP TRANSPORTASI UMUM MRT JAKARTA MENGGUNAKAN NAIVE BAYES CLASSIFIER

¹Diana Ikasari, ²Yuliana Fajarwati, ³Widiastuti

^{1,2,3}Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

¹yulianafajarwati@student.gunadarma.ac.id, ²d_ikasari@staff.gunadarma.ac.id,

³widiastuti@staff.gunadarma.ac.id

Abstrak

Penggunaan media sosial sebagai sarana untuk mengakses dan menyebarkan informasi telah banyak digunakan, salah satunya menggunakan media sosial Twitter. Twitter dalam penelitian ini digunakan sebagai sumber data untuk menganalisis tweet berbahasa Indonesia yang membahas mengenai transportasi MRT di Jakarta. Analisis sentimen pada twitter MRT Jakarta digunakan untuk melihat kecenderungan respon pengguna MRT Jakarta apakah berkecenderungan positif atau negatif berdasarkan hasil tweet dari Twitter MRT Jakarta. Analisis sentimen ini dapat membantu masyarakat Indonesia dalam menentukan pilihan transportasi umum yang nyaman dan aman berdasarkan ulasan transportasi umum dari Twitter oleh pengguna MRT Jakarta. Hasil penelitian ini dapat digunakan untuk meningkatkan sistem pada MRT Jakarta, baik dalam meningkatkan layanan maupun fasilitas agar menarik masyarakat untuk menggunakan MRT Jakarta sebagai alat transportasi. Analisis sentimen ini menggunakan metode Naïve Bayes Classifier yang merupakan metode pengklasifikasian. Tahap penelitian pertama yang dilakukan yaitu crawling, preprocessing yang terdiri dari case folding, cleansing, stopword removal, stemming, convert emoticon, dan tokenisasi. Tahap klasifikasi dilakukan setelah melalui fase preprocessing, dimana hasil klasifikasi tweet berkecenderungan positif atau negatif, menggunakan metode Naïve Bayes Classifier. Akurasi sistem pada analisis sentimen terhadap tweet yang terdapat dalam twitter MRT Jakarta adalah 95.88%.

Kata Kunci: Analisis Sentimen, MRT Jakarta, Naïve Bayes Classifier, Twitter.

Abstract

The use of social media as a means to access and disseminate information has been widely used, one of which is using social media Twitter. Twitter in this study is used as a source of information, in this case it is used as data to analyze Indonesian-language tweets discussing new public transportation MRT in Jakarta. Sentiment analysis on Jakarta MRT Twitter is used to see the tendency of the response of Jakarta MRT users whether positive or negative trends based on the results of tweets from Twitter MRT Jakarta. This sentiment analysis can help the people of Indonesia in making choices of convenient and safe public transportation based on public transportation reviews from Twitter by Jakarta MRT users. The results of this study can be used to improve the system on the Jakarta MRT, both in improving services and facilities to attract people to use the Jakarta MRT as a means of transportation. This sentiment analysis uses the Naïve Bayes Classifier method which is a classification method. The first step in the program is crawling, preprocessing which consists of case folding, cleansing, stopword removal, stemming, converting emoticons, and tokenization. The classification phase is carried out after going through a preprocessing phase, where the results of the classification of tweets have a positive or negative tendency, using the Naïve Bayes Classifier method. The accuracy of the system in sentiment analysis of tweets contained in the Jakarta MRT twitter is 95.88%.

Keywords: Sentiment Analysis, MRT Jakarta, Naïve Bayes Classifier, Twitter.

PENDAHULUAN

Kemacetan sering terjadi di berbagai wilayah, khususnya kota besar di Indonesia. Meningkat cepatnya pertumbuhan pergerakan orang dan barang, dimana permintaan penyediaan jaringan jalan yang akhirnya memunculkan kebutuhan transportasi. Kota menurut konteks perencanaan sistem transportasi regional maupun nasional mempunyai fungsi sebagai simpul jasa distribusi yang memiliki peran dominan dalam hal pemacu tingkat pertumbuhan ekonomi. Adanya perpindahan barang dan manusia yang semakin kompleks seiring dengan perkembangan kota mengakibatkan adanya sistem transportasi[1].

Peningkatan kualitas dan fasilitas transportasi umum akan membuat pengguna kendaraan pribadi beralih menggunakan transportasi umum, sehingga dapat mengurangi jumlah kendaraan yang membuat kemacetan. Kenyataannya masih banyak masyarakat enggan menggunakan transportasi umum dan lebih memilih menggunakan kendaraan pribadi dengan alasan beberapa faktor diantaranya mengenai waktu perjalanan, ongkos, keselamatan dan keamanan, serta kesenangan dan kenyamanan pengguna transportasi umum tersebut. Hal ini perlu diperhatikan lebih lanjut dengan mengetahui pendapat masyarakat tentang apa yang dirasakan saat menggunakan transportasi umum dalam kota sehingga bisa menjadi tolak ukur dan mampu memberikan beberapa

beberapa solusi untuk pengambilan keputusan guna meningkatkan mutu layanan hal lainnya.

Teknologi yang berkembang begitu cepat, mengakibatkan mudahnya untuk mengakses dan mendapatkan informasi, salah satunya adalah mengakses informasi melalui penggunaan sosial media. Berdasarkan penelitian [2] data yang diperoleh pada tahun 2015 menggambarkan bahwa pertumbuhan *twitter* sangat cepat, terlihat bahwa pengguna aktif internet di Indonesia mencapai angka 72.7 juta pengguna. Hal ini membuktikan bahwa media sosial sangat erat hubungannya dengan perkembangan teknologi. *Twitter* bisa dianggap sebagai perkembangan dari layanan pesan singkat atau SMS, dengan platform berbasis internet. *Twitter* merupakan utilitas yang memungkinkan pengguna mengirim SMS ke seluruh dunia, menjadikan pengguna dapat mengekspresikan pendapat, informasi, dan kegiatan sehari-hari. Berdasarkan data yang dirilis oleh *Twitter* Indonesia, bahwa Indonesia merupakan salah satu dari 5 negara terbesar dalam penggunaan *Twitter* secara aktif, yakni 79% dari pengguna *Twitter* Indonesia merupakan pengguna aktif [3]. Reaksi masyarakat Indonesia baik yang berbentuk opini maupun fakta terhadap berbagai hal yang terjadi, terkumpul di dalam media sosial *Twitter* dan muncul sebagai *tweets*. Isi *Tweets* ini menjadi sumber data apabila diolah dengan benar dan tepat dapat menghasilkan berbagai informasi yang berguna, misalnya dapat digunakan sebagai pengambilan keputusan atau digunakan untuk

meningkatkan persepsi seseorang terhadap suatu hal yang baru diketahui. Bentuk analisis pendapat atau *review* dapat dilakukan menggunakan analisis sentimen, di mana analisis sentimen merupakan proses yang dilakukan memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi [4]. Salah satunya dengan menganalisis sentimen masyarakat terhadap suatu hal dimana dalam penelitian ini menggunakan hasil *tweets* dari pengguna MRT Jakarta.

Beberapa penelitian terkait yang menjelaskan tentang analisis sentimen berdasarkan *tweet* telah dilakukan oleh Anurag P. Jain dan Vijay D. Kattar bahwa perkembangan media sosial seperti *Twitter*, di mana penggunaan data *Twitter* untuk analisis sentimen terhadap sebagian produk atau orang atau sebuah kegiatan (*event*) menggunakan *K-neighbour classifier* [5]. Penelitian yang dilakukan [6], membahas tentang bagaimana melakukan analisis sentimen terhadap opini dari berbagai jenis jaringan sosial seperti *Twitter*. Analisis sentimen dilakukan berdasarkan *tweet* yang telah diberi label positif atau negatif, menggunakan leksikal analisis melakukan proses membandingkan *tweet* yang di-*crawler* dengan *Bag of Word* (BOW) untuk mendefinisikan berkecenderungan positif atau berkecenderungan negatif. Penelitian lain yang terkait adalah penelitian yang dilakukan oleh [7], menjelaskan tentang aplikasi sosial yang semakin populer mem-*publish emotion* dalam bentuk

review yang digunakan untuk menefusikan analisis sentimen berdasarkan algoritma *Belief Maximasion* dengan *Supervised topic level Sentiment Model* (SSM).

Berbeda dengan sosial media lainnya, *Twitter* memiliki keterbukaan terhadap data yang dimilikinya melalui API (*Application Programming Interface*). Melalui API, *tweets* yang terdapat dalam *Twitter* dapat diakses sesuai dengan kebutuhan pengguna, baik *keyword* yang diinginkan maupun rentang waktu yang dibutuhkan sehingga informasi menjadi mudah didapatkan dan diolah menjadi suatu informasi yang berguna. *Tweets* yang dikumpulkan dan dianalisis disebut sebagai analisis sentimen. Menurut [8] pada berbagai penelitian yang mengambil tema analisis dokumen tekstual yang pernah dilakukan, metode Naive Bayes merupakan metode yang memiliki tingkat akurasi yang paling tinggi.

Penelitian [9] melakukan analisis sentimen terhadap dokumen teks yang menghasilkan akurasi sebesar 83%. Pada penelitian [10] melakukan pengujian terhadap beberapa algoritma yang biasa digunakan untuk mendeteksi *spam*, hasil akurasi yang didapatkan sebesar 96.5%. Proses seleksi fitur untuk mereduksi fitur yang tidak relevan dalam proses klasifikasi. Metode seleksi fitur yang umum digunakan adalah *Chi Square* (X^2), *Mutual Information* (MI), dan *Frequency-Based* [9], mengatakan untuk kasus seleksi fitur *frequency-based* memiliki kinerja yang buruk dibandingkan MI dan Chi

Square. Berdasarkan hal tersebut, langkah awal dalam proses pengembangan sistem adalah *spam filtering*, penelitian ini berusaha mengetahui performa algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan *tweet* berbahasa Indonesia ke dalam bentuk *tweet* negatif atau *tweet* positif. Indonesia, khususnya Jakarta memiliki transportasi umum baru yaitu MRT (*Mass Rapid Transit*) Jakarta. Adanya transportasi umum MRT ini dapat memberikan kontribusi dalam meningkatkan kapasitas transportasi umum. Tingkat kepuasan masyarakat Jakarta dengan adanya transportasi umum MRT Jakarta dapat menjadi tolak ukur untuk meningkatkan fasilitas menjadi lebih baik dan membuat pengguna menjadi aman dan nyaman. Analisis sentimen pada *Twitter* transportasi umum MRT Jakarta digunakan untuk melihat kecenderungan respon masyarakat pengguna MRT Jakarta apakah berkecenderungan positif atau negatif, berdasarkan hasil *tweet* dari *Twitter* dapat membantu masyarakat Indonesia agar lebih mudah memilih transportasi umum yang nyaman dan aman, bagi pihak manajemen MRT dapat membantu meningkatkan layanan dan fasilitas pada MRT Jakarta sehingga memberikan rasa nyaman, aman dan tepat waktu bagi pengguna. Berdasarkan latar belakang tersebut dilakukanlah penelitian mengenai analisis sentimen masyarakat Jakarta terhadap transportasi umum MRT Jakarta, dengan melihat *tweets* masyarakat melalui sosial

media *Twitter* berbahasa Indonesia menggunakan Metode *Naïve Bayes Classifier*.

METODE PENELITIAN

Penelitian dilakukan dengan dua tahapan yaitu *training* dan *testing* seperti terlihat pada Gambar 1. Pada tahap *training* dilakukan proses *crawling* (pengumpulan data) pada *tweet*, data latih diklasifikasikan secara manual ke dalam kelas sentimen positif dan negatif. Tahap *preprocessing* adalah *case folding*, *cleansing*, *stopword removal*, *stemming*, *convert emoticon*, dan *tokenisasi*. *Tweet* yang sudah melalui tahap *preprocessing*, dibuat model probabilistik yang akan digunakan pada tahap *testing* sebagai model analisis sentimen, sedangkan pada tahap *testing* menggunakan *tweet* data uji, tahap *testing* melalui *crawling* dan *preprocessing*. Data uji diproses sesuai dengan model analisis sentimen yang dibuat pada tahap *training* sehingga menghasilkan klasifikasi berupa sentimen positif atau negatif. Pada klasifikasi dilakukan perhitungan berdasarkan metode *Naïve Bayes* untuk menampilkan hasil klasifikasi berupa perhitungan angka, kecepatan dan akurasi.

Data penelitian ini merupakan kumpulan *tweet* yang diambil dari akun MRT Jakarta pada *Twitter* sebanyak 100 data. Data *tweet* yang digunakan sudah diberi sentimen positif atau negatif, dan sudah dikategorikan ke dalam data uji maupun data latih.

Crawling Data

Data yang diambil merupakan data *tweet* yang terdapat dalam *Twitter* menggunakan koneksi untuk mengakses *API Twitter*. Pengaksesan data *tweet* *Twitter* memerlukan hak akses untuk dapat mengakses data *tweet* berupa *consumer key*, *consumer secret*, *access token*, dan *access token secret*.

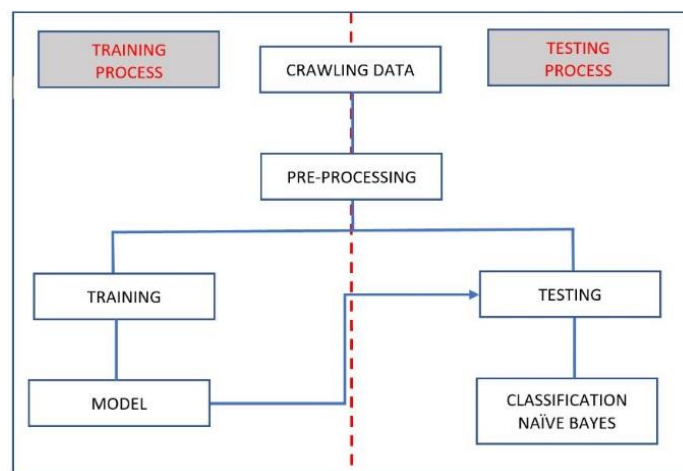
Data Pre-Processing

Pada tahap *preprocessing* dilakukan langkah-langkah yaitu *case folding*, *cleansing*, *stopword removal*, *stemming*, *convert emoticon*, dan *tokenizing*. Tahap *case folding* dilakukan untuk pengubah huruf pada *Twitter* menjadi huruf kecil. Hanya huruf “a” sampai

dengan “z” yang diterima dengan memeriksa ukuran setiap karakter dari awal sampai akhir karakter.

Apabila ditemukan karakter yang menggunakan huruf kapital (*uppercase*) maka huruf tersebut akan diubah menjadi huruf kecil (*lowercase*). Tahap *cleansing* dilakukan untuk menghilangkan *noise* (kesalahan acak atau varian dalam variabel terukur yang terdapat pada kalimat).

Kata yang dihilangkan dalam penelitian adalah karakter HTML, hastag (#), *username*(@username),*url*(http://website.co), dan *email*(nama@gmail.com). Tabel 1 memperlihatkan ilustrasi dari tahapan *cleansing*.



Gambar 1. Tahapan Penelitian

Tabel 1. Ilustrasi Tahapan *Cleansing*

Data Hasil <i>Case folding</i>		Tahapan <i>Cleansing</i>	
Input	rt @mrtjakarta: teman mrt perjalananmu termasuk kawasan ganjil genap? yuk naik mrt! mrt jakarta melewati kawasan ganjil genap dari lebak bulus.	Output	teman mrt perjalananmu termasuk kawasan ganjil genap? yuk naik mrt! mrt jakarta melewati kawasan ganjil genap dari lebak bulus.

Tahap selanjutnya adalah *stopword removal*. Daftar kata *stopword* pada penelitian bersumber dari penelitian [11], Pada tahap ini, kumpulan *tweet* melalui tahap *stopword removal*. Setiap kata pada *tweet* diperiksa. Apabila terdapat kata sambung, kata depan, kata ganti, atau kata yang tidak ada hubungannya dalam analisis sentimen, maka kata tersebut akan dihilangkan. Kata-kata yang terkandung pada daftar *stopword* yang terdapat pada daftar kata *stopword* Bahasa Indonesia berisi kata-kata yang sering muncul namun tidak memiliki arti. Pada contoh kata “bagaimana”, “juga”, “agar” dan “jadi” terdapat di tabel kata *stopword* sehingga kata tersebut harus dihilangkan.

Tahap *stemming* dilakukan untuk mengubah kata ke bentuk dasar dengan cara menghilangkan imbuhan-imbuhan pada kata dalam dokumen. Algoritma *stemming* yang digunakan dalam penelitian ini adalah Algoritma Nazief Adriani. Tahap *convert emoticon* dilakukan dalam penelitian adalah meng-*convert icon* sedih atau senang menjadi *text*, selain *emoticon* sedih dan senang akan dihapus atau diabaikan. *Tokenisasi* merupakan proses pemotongan *string* input berdasarkan kata yang menyusunnya serta membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan. Langkah-langkah pada tahap *tokenisasi* adalah menggunakan kata hasil dari *convert emoticon*, kemudian memotong setiap kata dalam teks berdasarkan pemisah kata seperti titik (.), koma (,), dan spasi dan bagian

yang hanya memiliki satu karakter non alphabet dan angka akan dihilangkan.

Proses *Training*

Proses *Training* menggunakan data latih yaitu *tweet* yang sudah diklasifikasikan, bentuk term yang didefinisikan berkecenderungan positif atau negatif. Sebelum melakukan klasifikasi dengan *Naive Bayes* maka sebelumnya harus diperhatikan yaitu Kosakata (*vocabulary*). Kosakata adalah jumlah kata (*term*) pada semua data latih yaitu *tweet* yang sudah diklasifikasikan. Data latih kemudian melalui tahap *preprocessing*. Data diambil dari hasil klasifikasi, tabel 2 merupakan contoh *tweet* data latih dengan klasifikasi manual, yang terdiri atas 7 data *tweet* berkecenderungan positif dan 3 data *tweet* berkecenderungan negatif. Berdasarkan data pada Tabel 2 dapat digunakan untuk menghitung *prior probability* dimana setiap fitur yang merepresentasikan pesan dihitung probabilitas kemunculannya di masing-masing kelasnya dengan menggunakan persamaan 1.

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (1)$$

Nilai prior probabilitas kelas positif data berdasarkan Tabel 2 adalah -0,1549 dan prior probabilitas negative bernilai -0,5228. Himpunan data latih pada Tabel 2 kemudian diproses dengan memecah kalimat sehingga menghasilkan kata atau *term*, kemudian menghitung kemunculan kata pada setiap sentimennya seperti dapat dilihat pada Tabel 3.

Banyaknya kosa kata yang dihasilkan dari pemecahan kalimat, maka dilakukan perhitungan setiap kata pada masing-masing kategori dengan menggunakan persamaan 2.

$$P(X_i|V_j) = \frac{n_k+1}{n+kosakata} \quad (2)$$

$\frac{\text{jumlah kemunculan kata}+1}{\text{jumlah kemunculan kata tiap kategori}+\text{jumlah semua}}$

Sebagai contoh pada Tabel 3 untuk kata “Mrt” dihitung *likelihood* menggunakan persamaan 2 dengan kemunculan kata dengan kelas positif sebanyak 2 dan negatif 0 didapat nilai *likelihood* positif sebesar -0,920818754 dan negatif sebesar -1,251811973. Nilai prior probabilitas dan *likelihood* yang dihasilkan untuk *term* “Mrt” pada proses training akan digunakan pada proses testing.

Tabel 2. Himpunan Data Latih

Data	Hasil Tokenisasi	Sentimen (V)
Tweet 1	Teman	Positif
Tweet 2	Mrt jalanmu	Positif
Tweet 3	Masuk	Positif
Tweet 4	Kawasan ganjil	Negatif
Tweet 5	Genap yuk naik	Positif
Tweet 6	Mrt Jakarta	Positif
Tweet 7	Lewat	Negatif
Tweet 8	Kawasan ganjil	Negatif
Tweet 9	Genap	Positif
Tweet 10	Lebak Bulus	Positif

Tabel 3. Kemunculan Kata pada Tiap Sentimen

Term	Jumlah Kemunculan Data Latih Positif (Vj)	Jumlah Kemunculan Data Latih Negatif (Vj)
Teman	1	0
Mrt	2	0
Jalanmu	1	0
Masuk	1	0
Kawasan	0	2
Ganjil	0	2
Genap	2	0
Yuk	1	0
Naik	1	0
Jakarta	1	0
Lewat	0	1
Lebak	1	0
Bulus	1	0
$\sum \text{kosakata} = 13$	$\sum n = 12$	$\sum n = 5$

Proses Testing

Pada proses *testing* dihitung nilai *posterior* probabilitas kelas menggunakan persamaan 3, dimana sebelumnya telah diketahui nilai prior probabilitas berdasarkan persamaan 2.

$$\text{Posterior Probabilitas kelas} = \frac{\text{prior probabilitas kelas} + \text{total log likelihood kelas}}{\text{total log likelihood kelas}} \quad (3)$$

Nilai *posterior* kelas positif adalah -1,017728754 dan *posterior* negatif sebesar -2,251811973. Selanjutnya dilakukan penentuan *log* terbesar dari *posterior* probabilitas kelas, berdasarkan perhitungan persamaan 3 diketahui nilai terbesar dari *posterior* probabilitas adalah -1,017728754. Sebagai contoh *term* yang digunakan adalah “Mrt”, menggunakan persamaan 4 dan 5, sehingga didapat nilai *posterior* positif sebesar 0,5761733 dan *posterior* negatif 0,4238266.

$$\text{Posterior Probabilitas Term Positif} = \frac{\text{positif probabilitas}}{\text{positif probabilitas} + \text{negatif probabilitas}}$$

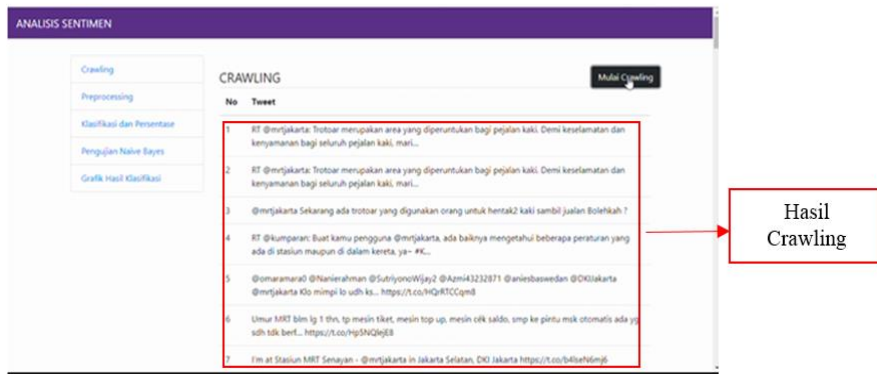
$$\text{Posterior Probabilitas Term Negatif} = \frac{\text{negatif probabilitas}}{\text{positif probabilitas} + \text{negatif probabilitas}} \quad (5)$$

Berdasarkan nilai *posterior* probabilitas nilai *posterior* positif memiliki nilai lebih besar dibandingkan dengan nilai negatif yaitu 0,5761733, dapat disimpulkan bahwa *term* “Mrt” pada data uji memiliki kecenderungan pada sentimen POSITIF.

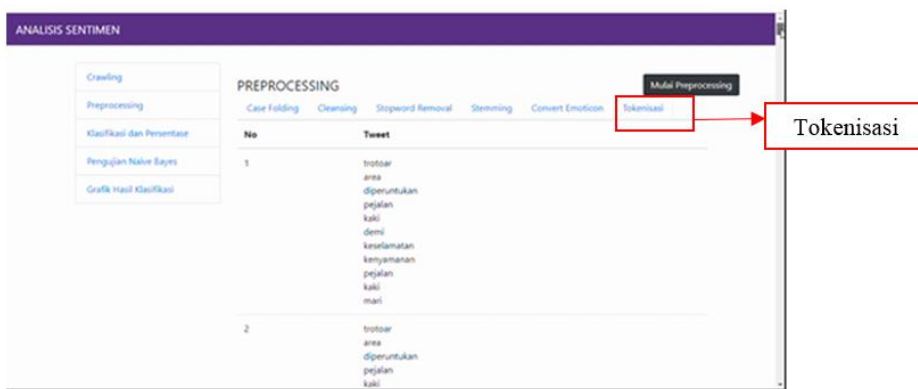
HASIL DAN PEMBAHASAN

Uji coba data terhadap aplikasi dimulai dengan proses *crawling* data *tweet* yang terlihat pada Gambar 2, data berupa *tweet* berisi opini masyarakat yang mengandung unsur transportasi umum MRT Jakarta sebanyak 100 *tweet* data terbaru.

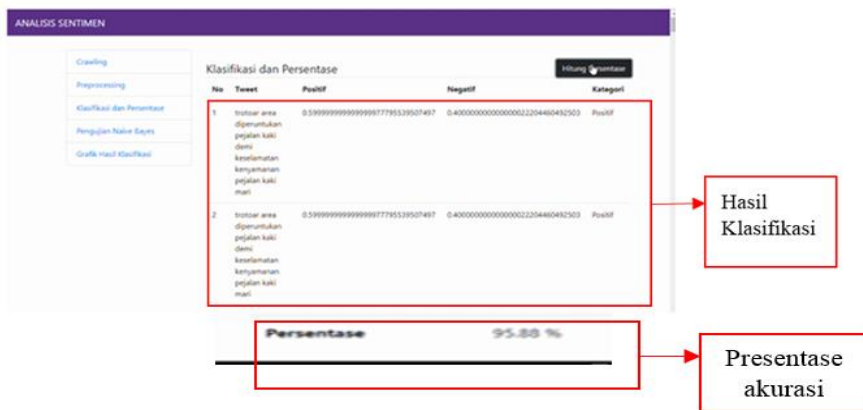
Data hasil *crawling* dilanjutkan ke tahap berikut yaitu *preprocessing*, terlihat pada Gambar 3. Tahap akhir yang dilakukan adalah proses klasifikasi untuk menentukan apakah data yang akan diuji termasuk ke dalam sentimen positif atau negatif. Pada tahap ini digunakan algoritma *Naive Bayes*, terlihat pada Gambar 4. Hasil klasifikasi dikategorikan berkecenderungan positif atau negatif berdasarkan perbandingan nilai *posterior* probabilitas, apabila perhitungan nilai *posterior* probabilitas positif lebih besar dari nilai *posterior* probabilitas negatif maka dihasilkan nilai klasifikasi dengan kecenderungan positif, begitu juga sebaliknya. (4)



Gambar 2. Tampilan Proses *Crawling* Data *Tweet*



Gambar 3. Tampilan Proses *Preprocessing*-Tokenisasi



Gambar 4. Tampilan Proses Klasifikasi dan Presentase Akurasi

Hasil Pengujian Perhitungan Klasifikasi

Pada pengujian aplikasi analisis sentimen digunakan 10 *tweet* dari 100 *tweet* pada *Twitter* akun MRT Jakarta dapat dilihat pada

Tabel 4, dihasilkan berdasarkan perhitungan pada persamaan 4 dan 5 untuk menentukan kecenderungan hasil klasifikasi positif atau negatif berdasarkan sistem.

Tabel 4. Hasil Pengujian Klasifikasi

No	Sentimen Asli	P Pos	P Neg	Hasil Analisis	Hasil
1	Positif	0.507640	0.492359	Positif	Akurat
2	Positif	0.507640	0.492359	Positif	Akurat
3	Positif	0.000456	0.000038	Negatif	Tidak Akurat
4	Negatif	0.002264	0.000137	Positif	Tidak Akurat
5	Negatif	0.000322	0.000677	Negatif	Akurat
6	Negatif	0.006482	0.008375	Negatif	Akurat
7	Positif	0.001659	0.000960	Positif	Akurat
8	Positif	0.003467	0.002058	Positif	Akurat
9	Negatif	0.000456	0.000690	Negatif	Akurat
10	Negatif	0.000215	0.000259	Negatif	Akurat

Tabel 5. Hasil Confusion Matrix

Jumlah Data Uji (10)	Sentimen Hasil Analisis	
	POSITIF	Sentimen Hasil Analisis NEGATIF
Sentimen Asli POSITIF	True Positives = 5	False Negatives = 2
Sentimen Asli NEGATIF	False Positives = 2	True Negatives = 1

Pengujian *Confusion Matrix*

Pengujian akurasi sistem dilakukan untuk mengetahui tingkat akurasi sistem untuk *tweet* yang dilakukan secara manual dengan klasifikasi *tweet* yang dilakukan oleh sistem dengan menggunakan Naive Bayes Classifier.

Pengujian dilakukan dengan menggunakan *confusion matrix* yaitu sebuah matrik dari prediksi yang akan dibandingkan dengan kelas yang asli dari data inputan. Pengujian dilakukan menggunakan 10 data uji yang diambil secara acak dan sudah diberi label. Data uji tersebut akan dibandingkan dengan hasil klasifikasi yang dilakukan oleh sistem. Tabel 5 adalah hasil dari *confusion matrix*. Setelah sistem melakukan klasifikasi, lalu dilakukan perhitungan akurasi berdasarkan Tabel 5, Akurasi sistem dihitung dengan membagi jumlah nilai *true* positif dan

true negatif dengan jumlah *true* positif, *false* negatif, *false* positif dan *true* negatif seperti pada persamaan 6.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

Sedangkan tingkat kesalahan sistem (*error rate*) menggunakan persamaan 7.

$$Error Rate = 1 - Akurasi \quad (7)$$

Data pengujian akurasi yang sebanyak 10 *tweet*, mendapatkan hasil akurasi 60%, tingkat kesalahan sistem 40%. Berdasarkan pengujian akurasi, didapatkan hasil akurasi klasifikasi *tweet* dari sistem analisis sentimen dengan menggunakan Naive Bayes Classifier sebesar 95.88% dengan *precision* positif sebesar 70% dan *precision* negatif sebesar 30%.

KESIMPULAN DAN SARAN

Analisis sentimen *tweet* dari *Twitter* akun MRT Jakarta (@mrtjakarta) telah berhasil dilakukan dengan menggunakan dua proses yaitu proses *training* dan proses *testing*. Sistem dapat memproses data dan mengklasifikasikan ke dalam sentimen positif atau negatif dengan menggunakan metode *Naive Bayes Classifier*. Pengukuran akurasi sistem dengan melakukan uji coba menggunakan *confusion matrix*. Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada analisis sentimen, dengan jumlah *tweet* data latih sebanyak 90 *tweet* dan data uji sebanyak 10 *tweet*, didapatkan akurasi sistem sebesar 95.88%. Percobaan berskala dan diambil rata-rata dengan data *training* dan data *testing* sebanyak 10 *tweet*. Hasil dari analisis sentimen dalam bentuk klasifikasi *tweet* yang memiliki nilai kecenderungan positif atau negatif dapat dimanfaatkan sebagai pengambilan keputusan dalam rangka meningkatkan layanan dan fasilitas bagi pengguna MRT oleh manajemen pengelola. Metode *Naive Bayes Classifier* dapat digunakan sebagai metode pengklasifikasian pada analisis sentimen karena tingkat akurasi yang besar.

Pengembangan sistem lebih lanjut dapat dilakukan pada penelitian berikutnya berupa penambahan jumlah data latih untuk mendapatkan hasil yang lebih akurat, menambahkan fitur perbandingan persentase positif

negatif dan memperindah tampilan *web* agar lebih menarik.

DAFTAR PUSTAKA

- [1] I. Wijanarko, M. A. Ridlo, "Faktor-faktor Pendorong Penyebab Terjadinya Kemacetan", *Jurnal Planologi*, Vol. 14, No.1, April 2017.
- [2] A. E. P. Herrera "Pemanfaatan Media sosial Twitter Oleh Ridwan Kamil dan Ganjar Pranowo Telah Sesuai dengan Fungsi Utama Media Massa", *The Messenger*, Vol III, No.2, Juli 2016.
- [3] Kemenkominfo. "Pengguna Internet di Indonesia 63 Juta Orang". from kominfo : https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker. 2019. [Accessed: Maret 2019].
- [4] B. Pang, L. Lee, *Opinion Mining and Sentiment Analysis*, Foundations and Trends in Information Retrieval, Vol. 2, No 1-2 pp. 1–135, 2008.
- [5] Anurag P. Jain, Vijay D. Katkar, "Sentimen Analysis of Twitter Data Using Data Mining", International Conference on Information Processing (ICIP), IEEE, 13 Juni 2016.
- [6] A. Pappu Rajan, S.P. Victor, "Web Sentiment Analysis for Scoring Positive or Negative Word Using Twitter Data", *International Journal of*

- Computer Application* (0975-8887), Vol. 96 No. 6. June 2014.
- [7] S. M. Basha, D. S. Rajput, “A Supervised Aspect Level Sentiment Model to Predict Overall Sentiment on Twitter Document”, *International Journal of Metadata Semantics and Ontologies*, Vol. 13 Issue 1, 4 Desember 2018.
- [8] Hairani, G. S. Nugraha, M. N. Abdillah, M. Innuddin, “Komparasi Akurasi Metode Correlated Naive Bayes Classifier dan Naive Bayes Classifier untuk Diagnosis Penyakit Diabetes”, *Jurnal Nasional Informatika dan Teknologi Jaringan*, Vol 3 No. 1, September 2018.
- [9] Juen Ling, I Putu Eka N. Kencana, Tjokorda Bagus Oka, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square”, *E-Jurnal Matematika* Vol. 3 (3), pp. 92-99, Agustus 2014.
- [10] Malarvizhi dan K. Saraswathi, “Web Content Mining Techniques Tools & Algorithms – A Comprehensive Study R “, *International Journal of Computer Trends and Technology (IJCTT)*, Vol 4 Issue 8, 2013.
- [11] Tala, Fadillah Z, *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*, Master of Logic Project Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands, 2003