

Analisis Sentimen Masyarakat Terhadap Kinerja Trans Semarang Menggunakan Metode *Random Forest*

Rizqi Fachreza^{1*}, Widiyanto Tri Handoko²

Teknik Informatika, Universitas Stikubank Semarang, Semarang, Indonesia

*e-mail *Corresponding Author*: rizqifachreza@mhs.unisbank.ac.id

Abstract

Opinion from the public on the performance of Trans Semarang transportation that has been widely discussed on social media X, resulting in many pros and cons. Sentiment analysis comes as a way to understand public opinion, examining the opinions and attitudes of individuals towards an object. individual attitudes towards an object. By using an algorithm methodology, namely Random Forest in classifying data to determine the accuracy of the data to determine the accuracy of the level of sentiment value of X users towards Trans Semarang performance. performance of Trans Semarang. This research uses a dataset taken with keyword "Trans Semarang". Random Forest algorithm is used to classify the data and then test it with various ratios, where the results of this algorithm can be used for the evaluation stage. The results of this algorithm can be used for the evaluation stage. This method produces a confusion matrix value with an accuracy of 81%, an average precision of 80%, average recall 80%, and average f-measure 80%.

Keyword: *Trans Semarang; Random Forest; Sentiment Analysis*

Abstrak

Opini dari masyarakat terhadap kinerja transportasi Trans Semarang yang ramai dibicarakan di sosial media X, sehingga menimbulkan banyak pro dan kontra. Analisis sentimen hadir sebagai cara untuk memahami opini publik, meneliti pendapat dan sikap individu terhadap suatu objek. Dengan menggunakan metodologi algoritma yaitu *Random Forest* dalam pengklasifikasian data untuk mengetahui akurasi tingkat nilai sentimen pengguna X terhadap kinerja Trans Semarang. Penelitian ini menggunakan *dataset* yang diambil dengan kata kunci "Trans Semarang". Algoritma *Random Forest* digunakan untuk mengklasifikasikan data dan kemudian mengujinya dengan berbagai rasio, dimana hasil algoritma ini dapat digunakan untuk tahap evaluasi. Metode ini menghasilkan nilai *confusion matrix* dengan *accuracy* 81%, *precision* rata-rata 80%, *recall* rata-rata 80%, dan *f-measure* rata-rata 80%.

Kata kunci: *Trans Semarang; Random Forest; Analisis Sentimen*

1. Pendahuluan

Transportasi merupakan solusi utama untuk mengangkut barang maupun manusia dari suatu tempat ke tempat yang diinginkan [1]. Untuk meminimalkan kemacetan lalu lintas di Kota Semarang masyarakat memilih Transportasi umum menjadi salah satu pilihan yang banyak diminati oleh warga Semarang saat ini contohnya Trans Semarang. Kinerja dari transportasi Trans Semarang ramai dibicarakan di dunia nyata maupun di dunia maya. Penilaian masyarakat terhadap kinerja Trans Semarang juga banyak dilakukan melalui media sosial, khususnya X yang menjadi *platform* yang paling efektif dan efisien. Dalam hal ini, tanggapan masyarakat terhadap kinerja Trans Semarang telah memunculkan berbagai hal mulai dari netral, positif, dan negatif. X merupakan salah satu platform media sosial populer yang mengandung berbagai berita, pemikiran, dan pandangan masyarakat secara real time [2]. Post merupakan ungkapan singkat pengguna dalam bentuk teks dengan batasan maksimal 280 karakter. Batasan ini mendorong pengguna untuk lebih ringkas dan ekspresif dalam menyampaikan idenya dibandingkan platform lain. Hal ini menjadikan post sebagai sumber data berharga untuk analisis sentimen, karena mengandung nilai emosi dan opini pengguna [3].

Dengan memahami persepsi publik terhadap Trans Semarang melalui media sosial X, maka diperlukan sebuah metode yang dapat lebih mudah mengidentifikasi area yang perlu diperbaiki dan membuat keputusan yang lebih baik untuk meningkatkan layanan mereka.

Analisis sentimen merupakan ilmu yang berguna untuk mengetahui opini masyarakat mengenai sesuatu yang terjadi, biasanya berbentuk komentar tulisan [4]. Analisis sentimen berguna untuk memahami opini publik, meneliti pendapat dan sikap individu terhadap suatu objek, seperti peristiwa, produk, ataupun isu tertentu. Dilakukan dengan menganalisis teks dari berbagai media sosial, seperti blog, X, dan *Facebook*, analisis sentimen membantu mengidentifikasi dan mengelompokkan opini publik ke dalam kategori positif, negatif, ataupun netral. Tujuannya untuk memahami persepsi publik terhadap suatu objek dan membantu pengambilan keputusan dalam berbagai bidang [5].

Dalam melakukan analisis sentimen, disini penulis menggunakan metode *Random Forest* sebagai algoritmanya. *Random Forest* adalah sebuah kumpulan metode pembelajaran yang menggunakan pohon keputusan sebagai base classifier yang dibangun dan dikombinasikan. Beberapa aspek penting dari metode ini termasuk penggunaan sampling terpandu untuk konstruksi pohon prediksi. Setiap pohon keputusan menggunakan prediktor acak, dan *Random Forest* melakukan prediksi dengan menggabungkan hasil dari setiap pohon keputusan melalui *majority vote* untuk klasifikasi dan rata-rata untuk regresi [6]. Penelitian ini bertujuan untuk menganalisis apa yang terjadi terhadap kinerja Trans Semarang dengan menggunakan metode *Random Forest*. Selain itu penelitian ini memiliki manfaat untuk manajemen Trans Semarang agar dapat mengambil keputusan terhadap masalah yang terjadi dengan hasil yang telah didapat dari analisis sentimen ini.

2. Tinjauan Pustaka

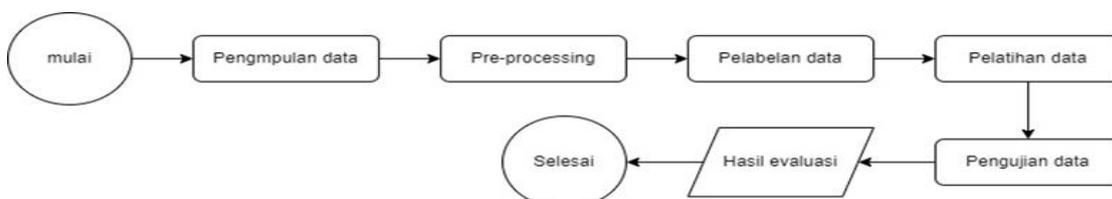
Dalam penelitian yang dilakukan oleh Nur Fitriyah, Budi Warsito, dan Di Asih I Maruddani pada tahun 2020 [7], penelitian ini membahas mengenai penggunaan metode *SVM* untuk menganalisis sentimen terhadap Gojek di media sosial Twitter. Penelitian ini bertujuan untuk mengklasifikasikan respon pengguna Gojek kedalam dua sentimen, yaitu sentimen positif dan sentimen negatif. Penelitian ini menggunakan algoritma *SVM* dengan melibatkan dataset berupa postingan di media sosial Twitter. Penelitian ini berhasil mendapatkan akurasi senilai 79,19%.

Penelitian yang dilakukan oleh Brata Mas Pintoko dan Kemas Muslim 2018 [5], memaparkan tentang sentimen masyarakat terhadap jasa transportasi online dengan penggunaan metode *Naïve Bayes Classifier*. Penelitian ini memiliki tujuan untuk mengetahui pandangan masyarakat terhadap penggunaan jasa transportasi online. Metode yang digunakan oleh peneliti adalah *Naïve Bayes* dengan melibatkan total 2000 data. Data yang diambil berasal dari Twitter. Dengan menggunakan metode *Naïve Bayes*, peneliti memperoleh hasil akurasi 88,60%.

Penelitian ini membahas mengenai analisis sentimen terhadap pengguna Gojek dan Grab pada media sosial Twitter menggunakan *Random Forest* yang dilakukan oleh Melia, Bambang Irawan, Odi Nurdiawan pada tahun 2023 [8]. Tujuan dilakukannya penelitian ini yaitu untuk menyelesaikan masalah antara klasifikasi dengan regresi menggunakan metode *Random Forest*. Penelitian ini mendapat nilai akurasi sebesar 76,25%.

Berdasarkan beberapa penelitian yang telah dilakukan, secara keseluruhan penelitian tersebut mengarah pada analisis sentimen terhadap jasa transportasi dengan penggunaan beberapa metode untuk mengklasifikasikan sesuatu kedalam beberapa kelompok. Dalam penelitian ini, yang membedakan antara penulis dengan penelitian sebelumnya yaitu penggunaan algoritma yang berbeda serta objek penelitiannya yaitu berobjek pada media sosial X.

3. Metodologi



Gambar 1. Desain Penelitian

Secara keseluruhan penelitian ini memiliki beberapa tahapan yaitu pengumpulan data, *Pre-Processing* data, pelabelan data, pelatihan data, pengujian model, evaluasi hasil. Proses ini dilakukan untuk mengevaluasi hasil dari model *Random Forest*.

3.1 Pengumpulan data

Menurut [9], crawling data merupakan metode yang digunakan untuk mengumpulkan atau mengunduh informasi dari kumpulan data. Pada awalnya metode crawler digunakan oleh browser internet untuk mengisi daftar mereka. Masalah yang dihadapi dalam menyelesaikan prosedur pengumpulan mencakup struktur informasi yang sangat besar dan kurangnya kontrol yang terkonsentrasi terhadap konten [10].

Data yang digunakan dalam penelitian ini adalah post dari pengguna terhadap kinerja Trans Semarang yang diambil dari media sosial X. Data ini mencakup teks opini yang diberikan oleh pengguna. Proses *crawling data* dilakukan dengan memanfaatkan modul *python* yang telah tersedia yaitu *Tweet Harvest*. Data yang telah diambil kemudian disimpan dalam format file CSV.

3.2 Pre-Processing data

Pre-Processing merupakan proses menghilangkan permasalahan menjadi data yang bersih agar dapat diproses selanjutnya [11]. *Pre-Processing data* merupakan proses dimana dilakukan pembersihan dan pemilihan data untuk mendapat data yang bersih dan siap untuk digunakan. *Pre-Processing* dilakukan untuk mengurangi data yang error atau noise. *Pre-Processing data* terdapat beberapa langkah yaitu ,sebagai berikut :

Cleaning data merupakan proses pembersihan data dari karakter yang diperlukan seperti tanda baca, emoticon, username, hashtag, alamat website, dan sebagainya. Selain itu, *cleaning data* juga berguna untuk menghapus data yang kosong maupun data yang duplikat. Proses *cleaning* dilakukan agar data lebih mudah dipahami.

Case folding, pada proses ini data yang telah dibersihkan dilakukan penyamaan semua huruf menjadi huruf kecil semua.

Normalisasi merupakan teknik yang digunakan untuk mengubah teks yang sulit dimengerti menjadi teks yang baku agar memenuhi tujuan tertentu [12].

Tokenisasi, proses menghilangkan tanda baca, seperti simbol ataupun karakter khusus lainnya pada teks [13]. Pada tahapan ini, dilakukan pemisahan sebuah kalimat menjadi beberapa kata/token. Pada tahapan ini juga dilakukan pengecekan pada karakter pertama hingga karakter terakhir, apabila karakter pertama bukan karakter pemisah seperti titik, koma, spasi, dan karakter pemisah lainnya.

Stemming, pada tahap ini merupakan tahap terakhir pada *Pre-Processing*, ini dilakukan untuk mengubah kata berimbuhan menjadi kata dasarnya. Kata imbuhan yang dihilangkan terdiri dari awalan, akhiran, sisipan, dan gabungan awal-akhiran.

3.3 Pelabelan data

Pelabelan data dilakukan berdasarkan post yang diunggah oleh pengguna di media sosial X. Pelabelan dilakukan secara manual, dengan cara memberikan polaritas secara logika peneliti terhadap kalimat opini. Kalimat opini akan diklasifikasikan kedalam 3 polaritas, yaitu positif, negatif, dan netral.

3.4 Pembentukan fitur

Data hasil dari *Pre-Processing* yang berupa kalimat akan diubah menjadi bentuk angka dengan dilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur. Salah satu metode yang digunakan untuk pembentukan fitur yaitu dengan menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*). *TF-IDF* yaitu metode penggabungan dua konsep antara frekuensi kemunculan sebuah kata dalam satu dokumen dan *inverse* frekuensi dokumen yang mengandung kata tersebut [14].

3.5 Pemodelan *Random Forest*

Random Forest merupakan algoritma supervised learning yang menggunakan banyak pohon putusan sebagai base classifier yang dikombinasikan sehingga menghasilkan suatu putusan [15]. Tidak ada rumus matematika yang dapat digunakan untuk menghitung *Random*

Forest secara keseluruhan. Algoritma ini menghasilkan prediksi yang lebih akurat dan stabil dengan menggabungkan banyak pohon keputusan individual.

Berikut ada beberapa rumus yang digunakan secara mendasar untuk *Random Forest*:

Menentukan *Entropy*

Entropy mengukur ketidakpastian atau keragaman dalam suatu kumpulan data. Semakin tinggi nilai *Entropy*, semakin beragam data tersebut.

$$Entropy(S) = \sum_{i=1}^c p_i \log_2(p_i) \tag{1}$$

Perhitungan *Entropy* untuk subset data

Data dibagi menjadi beberapa subset tergantung pemilihan kata kunci.

$$Entropy(S_i) = \sum_{i=1}^c p_i \log_2(p_i) \tag{2}$$

Menentukan *Weight Entropy* dari subset

$$Entropy(T) = \frac{S_1}{n_{total}} Entropy(S_1) + \frac{S_2}{n_{total}} Entropy(S_2) \tag{3}$$

Menentukan *Gain Information*

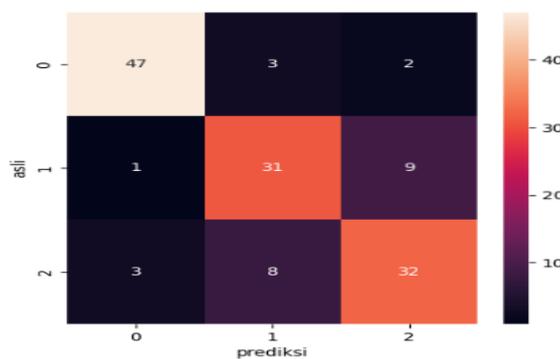
$$Gain = Entropy(S) - Entropy(T) \tag{4}$$

Dalam *Random Forest*, proses ini diulang berkali-kali untuk berbagai subset data dan fitur. Setiap pohon keputusan dibuat menggunakan subset acak dari data pelatihan dan subset acak dari fitur. Keputusan akhir ditentukan oleh agregasi hasil dari semua pohon.

Karena data yang digunakan pada penelitian ini cukup banyak dan rumit, maka digunakan alat bantu pada bahasa pemrograman *Python* dengan *library scikit-learn*.

3.6 Hasil Evaluasi

Pada penelitian ini penulis melatih sejumlah 100 buah model *Decision Tree* dan rasio 90:10 menghasilkan akurasi yang cukup bagus yaitu mencapai 81%. Dalam pengujian ini digunakan 136 data sebagai data uji. Hasil klasifikasi divisualisasikan dalam bentuk *confusion matrix*. Pada *confusion matrix* menghasilkan nilai *accuracy*, *precision*, *recall*, *F-measure*. *Confusion matrix* disajikan dalam bentuk gambar berikut:



Gambar 2. Visualisasi *confusion matrix*

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Seluruh Prediksi}} \tag{5}$$

$$Precision = \frac{\text{Jumlah Prediksi Benar (positif/netral/negatif)}}{\text{Jumlah Prediksi (positif/netral/negatif)}} \tag{6}$$

$$Recall = \frac{\text{Jumlah Prediksi Benar (positif/netral/negatif)}}{\text{Jumlah Asli (positif/netral/negatif)}} \tag{7}$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (8)$$

4. Hasil dan Pembahasan

4.1 Pengumpulan data

Mengumpulkan data dengan cara crawling data dari platform X adalah langkah pertama dalam melakukan penelitian ini. Data dikumpulkan menggunakan program python di Google Colabs. Data dapat dikumpulkan setiap 500 post dalam 20 menit, jadi butuh waktu yang cukup lama untuk mendapatkan data ini. Penulis dapat mengumpulkan 955 data dalam penelitian ini dalam waktu kurang lebih 40 menit. Dimulai dengan penginstalan library pandas, node.js, dan npm pada Google Colab.

	conversation_id_str	created_at	favorite_count	full_text	id_str
0	1786636690339361193	Sat May 04 06:24:09 +0000 2024	0	@yexoliny tidak kak karena kagok arah turun ha...	1786642956155510998
1	1786636690339361193	Sat May 04 06:20:30 +0000 2024	0	@Transsemarang di kagok minn	1786642037108011384

Gambar 3. Dataset Trans Semarang

4.2 Penginstalan library

Sebelum masuk ke tahap *Pre-Processing data*, maka perlu adanya penginstalan beberapa *library* yang dibutuhkan untuk masuk kedalam proses pengolahan data. *Library* yang dibutuhkan yaitu *emoji*, *Sastrawi*, *tweet-preprocessor*, *pandas*, *os*, *regular expressions*, *nltk*, *numpy*, *seaborn*, *sklearn* dan *matplotlib*.

4.3 Pre-Processing

Setelah dilakukan membaca data, data tidak dapat diolah secara mentah, harus ada proses pembersihan terlebih dahulu pada data tersebut. Yang perlu dilakukan selanjutnya yaitu tahap *Pre-Processing data*. *Pre-Processing* merupakan tahap yang digunakan untuk proses pembersihan sebuah dataset dari kata yang tidak berbobot. Langkah awal dalam Berikut beberapa tahap *Pre-Processing data* :

4.3.1 Cleaning data

Proses untuk membersihkan data dari karakter yang diperlukan seperti tanda baca, emoticon, username, hashtag, alamat website, dan sebagainya. Selain itu cleaning data juga berguna untuk menghapus data yang sama atau terduplikasi dan juga menghapus data yang kosong.

Tabel 1. Contoh hasil *cleaning data*

Sebelum Cleaning	Sesudah Cleaning
@yexoliny tidak kak karena kagok arah turun halte transit point kak~~	tidak kak karena kagok arah turun halte transit point kak
@terasipedas @yusufgunawan Trans semarang ok sih pelayanannya armadanya juga bnyk even yg di desa2 jadi ga perlu nunggu lama. Minus banyak supirnya yg ugal2an.	trans semarang ok sih pelayanannya armadanya juga bnyk even yg di desa jadi ga perlu nunggu lama minus banyak supirnya yg ugalan

4.3.2 Case Folding

Proses untuk membuat seluruh huruf dalam data menjadi huruf kecil ini bertujuan untuk mempermudah pada proses selanjutnya. Dengan script pada python yang telah dibuat maka dapat dilihat hasil berikut ini:

Tabel 2. Contoh hasil *case folding*

Sebelum	Sesudah
Perbandingan sama Trans Jakarta. Sayangnya saya ga dapet data untuk Trans Semarang soalnya statusnya bukan BUMD (status Trans Semarang adalah BLUD)	perbandingan sama trans jakarta. sayangnya saya ga dapet data untuk trans semarang soalnya statusnya bukan bumd (status trans semarang adalah blud)
Trans semarang ok sih pelayanannya armadanya juga bnyk even yg di desa2 jadi ga perlu nunggu lama. Minus banyak supirnya yg ugal2an.	trans semarang ok sih pelayanannya armadanya juga bnyk even yg di desa2 jadi ga perlu nunggu lama. minus. banyak supirnya yg ugal2an.

4.3.3 Normalisasi

Normalisasi merupakan langkah dimana suatu kata dikembalikan menjadi kata yang menggunakan ejaan yang baik. Contoh kata “bnyk” setelah di normalisasi menjadi “banyak”. Dan berikut adalah hasil dari setelah proses normalisasi :

Tabel 3. Contoh hasil normalisasi

Sebelum	Sesudah
Trans semarang ok sih pelayanannya armadanya juga bnyk even yg di desa2 jadi ga perlu nunggu lama. Minus banyak supirnya yg ugal2an.	trans semarang ok sih pelayanannya armadanya juga banyak even yg di desa2 jadi ga perlu nunggu lama. minus. banyak supirnya yg ugal2an.

4.3.4 Tokenisasi

Tahapan untuk memecah sebuah kalimat menjadi beberapa beberapa kata atau token untuk mengetahui kata aslinya. Ini adalah hasil dari proses tokenisasi :

```

0      [tidak, kak, karena, kagok, arah, turun, halte...
1      [di, kagok, min]
2      [turun, nya, di, halte, mana, dulu, kak, karen...
3      [min, kalo, turun, dr, koridor, terus, transit...
4      [sama, sama, kak]
...
1005  [jika, dari, unnes, naik, armada, koridor, ara...
1006  [mohon, maaf, kak, armada, trans, semarang, su...
1007  [terima, kasih, atas, apresiasinya, kak]
1008  [hallo, kak, armada, terakhir, dari, masing, m...
1009  [min, infokan, rute, unnes, the, park, mall, s...
Name: full_text, Length: 995, dtype: object
    
```

Gambar 4. Hasil Tokenisasi

Gambar tersebut adalah hasil dari tokenisasi dimana seperti contoh kalimat “di kagok min” dipisah per kata atau token menjadi [di, kagok, min].

4.3.5 Stemming

Pada tahap ini dilakukan untuk mencari kata dasar dari sebuah kata dengan menggunakan *function stemmer* dari *library Sastrawi*. Dengan kata lain, kata-kata yang ada dalam dataset dari tahap sebelumnya diubah menjadi kata dasar menurut KBBI.

```

tidak kak karena kagok arah turun halte transit point kak
di kagok min
turun nya di halte mana dulu kak karena tidak semua halte bisa untuk halte transit
min kalo turun dr koridor terus transit a itu bayar lagi ya
sama sama kak
    
```

Gambar 7. Hasil Stemming

Gambar diatas hasil perubahan kata-kata yang tidak baku menjadi baku seperti contoh pada tabel berikut ini :

Tabel 4. Contoh hasil *stemming*

Sebelum	Sesudah
di kagok minn	di kagok min

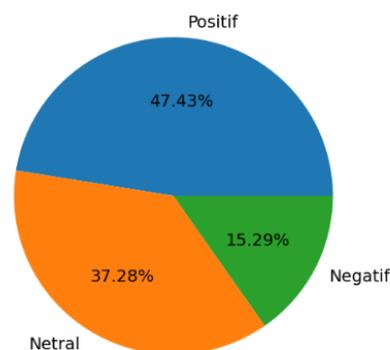
4.4. Pelabelan

Setelah proses *Pre-Processing* selesai, akan dilakukan pelabelan manual untuk mengetahui apakah sentimen dari masing-masing post bernilai positif, negatif, atau netral. Proses ini dilakukan dengan menggunakan Microsoft Excel, kemudian menyimpan file dalam format CSV. Berikut tampilan pelabelan manual :

1	full_text	sentimen
2	tidak kak karena kagok arah turun halte transit point kak	Positif
3	di kagok min	Netral
4	turun nya di halte mana dulu kak karena tidak semua halte bisa untuk halte transit	Netral
5	min kalo turun dr koridor terus transit a itu bayar lagi ya	Netral
6	sama sama kak	Positif
7	terima kasih min	Positif
8	stasiun semarang tawang sampai bawen only ribu untuk umum buruh rupiah saja ini rute panjang sptnya utk	Positif
9	naik armada trans semarang dari halte st tawang halte dekat dari kota lama naik armada koridoe arah cang	Netral
10	min info rute dari kota lama ke simpang lima	Netral
11	banding sama trans jakarta sayang saya ga dapet data untuk trans semarang soal status bukan bumd status	Negatif
12	ac trans semarang kalau siang g dingin beda sama trans jakarta kayak semua unit perlu di cek freon	Negatif
13	trackingnya bs andal transjakarta aja kalah sama trans semarang transjakarta mana bisa ada layan kyk gni t	Positif
14	gw yang dlu kuliah di semarang dan sekarang tinggal di bandung ketika kangen dengan trans semarang ini se	Positif
15	akan kami info lebih lanjut kak kait alih rute	Netral
16	dari halte batta pemuda naik armada koridor arah balaikota transit di halte balaikota dari halte balaikota ne	Netral
17	min jam sore nanti brt koridor masih lewat balaikota ga ya	Netral
18	siaapp terima kasih ya kak admin	Positif
19	jadi kangen jalan tidak jelas di semarang dari rumah mbah muter dulu di kota lama padahal halte paling deki	Negatif
20	sering naik transjakarta terus naik trans semarang rasa jauh banget beda bus jelek knalpot sering ngeluarin a	Negatif
21	selamat ulang tahun walikota semarang dr ir hj hevearita g rahayu msos moga panjang umur beri sehat dan	Positif
22	trans semarang ok sih layan armada juga bnyk even yang di desa jadi ga perlu nunggu lama minus banyak suj	Negatif
23	hallo kak dari pandanaran naik armada koridor arah mangkang bisa turun di halte balaikota halte dekat dari	Netral
24	dari terminal gunungpati naik armada koridor transit di halte balaikota dari halte balaikota naik armada kori	Netral
25	jog info dong trans semarang bisa bayar qris ga	Netral
26	jateng transportasi umum yang bagus wilayah semarang karo solo tapi di beberapa wilayah jateng udah ada	Positif

Gambar 5. Pelabelan Manual

Terdapat juga distribusi sentimen pada dataset Trans Semarang ,berikut adalah gambarnya :

Polaritas Sentimen pada Data X
(total = 955 post)

Gambar 6. Distribusi Sentimen Menggunakan Diagram Lingkara

Diagram diatas menjelaskan bahwa terdapat 47,43% pendapat positif Masyarakat terhadap kinerja Trans Semarang, 37,28% untuk pendapat netral, dan 15,29% untuk pendapat negative.


```

X_train , X_test , y_train , y_test = train_test_split(x_sm,y_sm, test_size=0.1,random_state=0)

random_forest_classifier = RandomForestClassifier(n_estimators=100,random_state=0)
random_forest_classifier.fit(X_train,y_train)

RandomForestClassifier
RandomForestClassifier(random_state=0)

random_forest_classifier_prediction = random_forest_classifier.predict(X_test)

```

Gambar 8. Proses membuat arsitektur dengan menggunakan library *sklearn*

Proses ini dimulai dengan membuat arsitektur *Random Forest* yang dapat diimport dari *sklearn*. Kemudian, model yang telah diimport dilatih dengan data latih. Setelah itu, data latih dapat digunakan untuk mengklasifikasi data uji. Hasil prediksi dari 136 data uji disimpan dengan nama "random_forest_classifier_prediction". Pada kasus ini penulis melatih sejumlah 100 buah model Decision Tree dan rasio 90:10, menghasilkan akurasi yang cukup bagus yaitu mencapai 81%.

0.8088235294117647

Gambar 9. Hasil Akurasi metode *Random Fores*

Terdapat juga beberapa hasil akurasi yang dilakukan dengan beberapa rasio yang berbeda, berikut adalah tabelnya :

Tabel 5. Uji coba beberapa data latih dan uji

Data Latih	Data Uji	Akurasi
90%	10%	0.8088235294117647
80%	20%	0.8051470588235294
70%	30%	0.7745098039215687
60%	40%	0.7794117647058824

4.7 Confusion Matrix

Setelah melakukan klasifikasi, langkah berikutnya adalah menguji model untuk memastikan efektivitas metode. Dalam pengujian ini, model klasifikasi data digunakan untuk mengukur perhitungan prediksi positif, negatif, dan netral. Dalam pengujian ini digunakan 136 data sebagai data uji. Hasil klasifikasi divisualisasikan dalam bentuk confusion matrix. Berikut juga terdapat hasil yang menunjukkan nilai *accuracy*, *precision*, *recall*, *F-measure*.

Tabel 6. Hasil report klasifikasi

Algoritma	Accuracy	Precision	Recall	F-measure
Random Forest	81%	80%	80%	80%

Dapat dilihat table diatas, bahwa hasil akurasi dari pelatihan data training sebanyak 136 data menghasilkan nilai 81%. Lalu untuk *precision* yaitu rasio prediksi benar positif dibanding dengan prediksi positif menghasilkan nilai 80%. Ada juga untuk *Recall* yang merupakan prediksi benar positif dengan keseluruhan data benar positif menghasilkan nilai 80%. Lalu untuk *F-measure* perbandingan rata-rata *precision* dan *recall* yang dibobotkan menghasilkan nilai 80%.

4.8 Pembahasan

Hasil penelitian ini digunakan untuk mengetahui dan menganalisis kinerja Trans Semarang berdasarkan komentar dari postingan sosial media X. Berdasarkan akurasi yang dihasilkan melalui metode *Random Forest* dapat mengukur perhitungan prediksi polaritas yaitu sebesar 81%.

Hasil penelitian ini juga didukung oleh Melia, Budi Irawan, dan Odi Nurdiawan (2023) [8], yang menggunakan model *Random Forest* untuk mengklasifikasikan sentimen di media sosial Twitter terkait Gojek dan Grab menghasilkan akurasi yang cukup tinggi yaitu 76,25%. Analisis sentimen terhadap aplikasi ruangguru menggunakan algoritma *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* yang dilakukan oleh Evita Fitri, Yuri Yuliani, Susy Rosyida, dan Windu Gata [15], penggunaan metode *Random Forest* dalam menganalisis sentimen menghasilkan akurasi yang cukup baik yaitu 97,16%. Temuan ini memberikan implikasi bahwa penggunaan metode *Random Forest* untuk menganalisis sentimen menghasilkan akurasi yang baik.

Berdasarkan hasil analisis sentimen yang telah dilakukan, penelitian ini menghasilkan sentimen yang positif terhadap kinerja Trans Semarang. Namun ada beberapa sentimen negatif yang muncul seperti ac bus yang tidak menyala, sopir bus yang mengendarai secara ugal-ugalan, dan juga ada efek asap dari knalpot bus yang dikeluarkan. Peneliti dapat memberi saran agar beberapa sentimen negatif tersebut dapat ditinjau oleh manajemen Trans Semarang sebagai bentuk perbaikan dimasa mendatang.

5. Simpulan

Penelitian ini menemukan bahwa platform X, yang memungkinkan pengguna untuk berbagi opini dan informasi secara publik, menyediakan data opini yang kaya untuk analisis sentimen. Dari data yang dianalisis, terdapat 47,43% komentar positif, 37,28% komentar netral, dan 15,29% komentar negatif. Algoritma *Random Forest* digunakan untuk mengklasifikasikan data dan menunjukkan performa yang baik dengan akurasi 81%, presisi rata-rata 80%, recall rata-rata 80%, dan f-measure rata-rata 80%. Hal ini menunjukkan bahwa platform X dan metode analisis sentimen yang digunakan dapat menjadi alat yang berharga untuk memahami opini publik.

Daftar Referensi

- [1] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi dan Sistem Informasi*, vol. 4, no. 1, pp. 96–106, 2023, DOI: 10.33365/jtsi.v4i1.2445.
- [2] H. Faradian, A. Rubhasy, and Y. F. Wijaya, "Analisis Sentimen Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naïve Bayes Classifier Pada Media Sosial X," *Jurnal Ilmiah Sain dan Teknologi*, vol. 2, no. 4, pp. 150–163, 2024.
- [3] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter," *Fountain of Informatics Journal*, vol. 4, no. 2, p. 41, Nov. 2019. DOI: 10.21111/fij.v4i2.3573.
- [4] B. Ramadhani, R. R. Suryono, and K. Kunci, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 714-725, 2024. DOI: 10.30865/mib.v8i2.7458.
- [5] B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," *e-proceeding of engineering*, vol. 5, no. 3, pp. 8121-8130, 2018.
- [6] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, pp. 122-130, 2022.
- [7] N. Fitriyah, B. Warsito, D. Asih, and I. Maruddani, "Analisis Sentimen Gojek pada Media Sosial Twitter dengan Klasifikasi Support Vector Machine (SVM)," *Jurnal Gaussian*, vol. 9, no. 3, pp. 376–390, 2020.
- [8] B. Irawan and O. Nurdiawan, "Analisis Sentimen Terhadap Pengguna Gojek dan Grab pada Media Sosial Twitter Menggunakan Random Forest," *JATI : Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 5, pp. 3614-3618, 2023.
- [9] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," in *Indonesian Symposium on Computing (Indo-SC)*, Bandung, 11–16, 24 September, 2016.
- [10] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling (Requirements Identification for NTB

- People in pandemic covid-19 at Social Media Using Crawling Method),” *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 2, no. 4, pp. 244–250, 2021.
- [11] F. Nurona Cahya *et al.*, “Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN),” *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, pp. 618–626, 2023.
- [12] R. Riyaddulloh and A. Romadhony, “Normalisasi Teks Bahasa Indonesia Berbasis Kamus Slang Studi Kasus: Tweet Produk Gadget Pada Twitter,” *e-proceeding of engineering*, vol. 8, no. 4, pp. 4216-4228, 2021.
- [13] L. Hermawati, V. Berland, A. Rahmadiyah, E. Hutabarat, and D. Dwi Saputra, “Komparasi Metode Text Mining Terhadap Masalah Pengklasifikasian Narasi Informative & Non Informative Pada twitter @PLN_123,” *Jurnal Sistim Informasi dan Teknologi*, vol. 5, no. 1, pp. 109-120, 2023. DOI: 10.37034/jsisfotek.v4i2.191.
- [14] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, “Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 306–312, 2018.
- [15] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *TRANSFORMTIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: www.nusamandiri.ac.id,