

TWITTER SENTIMEN GOJEK INDONESIA DAN GRAB INDONESIA

Hakiki Sandhika Raja^{1*}, Stephani Magdhalena¹

¹Badan Pusat Statistik Kabupaten Kutai Timur, Indonesia

Corresponding author: shandikaraja@bps.go.id

Abstrak. Twitter merupakan media yang memiliki karakteristik utama *text base marketing* menggunakan narasi pendek untuk mendapatkan *engagement* dari masyarakat. Makalah ini membahas bentuk *engagement* dalam bentuk respon yang didapatkan dari kumpulan *tweet* dua perusahaan transportasi daring di Indonesia yaitu Gojek dan Grab Indonesia. Penilaian respon pada penelitian ini menggunakan analisis sentimen yang memanfaatkan 3 metode klasifikasi yaitu *Random Forest*, *Support Vector Machine*, dan *Multinomial Naïve Bayes*. Hasil pengolahan menunjukkan bahwa metode *Support Vector Machine* memberikan nilai akurasi tertinggi dengan 76%. Perbandingan kepuasan yang digunakan adalah *Net Sentimen Score* (NSS) dimana persentase dari selisih *tweet* negatif dan *tweet* positif menjadi indikator yang menentukan kepuasan konsumen.

Kata Kunci: *twitter, analisis sentiment, transportasi daring.*

1 PENDAHULUAN

Transportasi daring saat ini sudah mulai menjadi kebutuhan hidup masyarakat, khususnya yang tinggal di daerah perkotaan. Hal ini secara langsung membuat aplikasi transportasi daring ramai digunakan oleh masyarakat, terlihat dari jumlah unduhan aplikasi 2 perusahaan besar transportasi daring di Indonesia yaitu Gojek dan Grab yang sudah melebihi 50 juta unduhan di Google *Playstore*. Dengan besarnya jumlah pengguna produk tersebut akan berbanding lurus karakteristik konsumen yang semakin beragam serta persaingan dalam mendapatkan konsumen baru dan cenderung loyal dalam menggunakan produk-produk yang ditawarkan oleh perusahaan. Salah satu cara untuk terus menjaga kepercayaan serta kenyamanan konsumen dan calon konsumen adalah *marketing strategy* yang *creative* dan mudah untuk dikonsumsi oleh masyarakat. Impresi terhadap komunikasi perusahaan akan menjadi faktor yang dapat digunakan untuk meningkatkan penjualan, hal ini sesuai dengan hasil penelitian yang dilakukan oleh *Pollay & Mital* pada tahun 1993 yang menyatakan “Iklan mempromosikan pesan sosial dan gaya hidup dengan menggambarkan posisi konsumen ideal dan merangsang keinginan konsumen untuk membeli^[1]”

Penggunaan sosial media merupakan salah satu alternatif untuk media *creative marketing* yang dilakukan oleh perusahaan transportasi daring, hal ini sejalan dengan karakteristik utama pengguna transportasi dari yaitu “melek teknologi”. Saat ini terdapat beberapa platform sosial media yang sering digunakan di Indonesia, yaitu Facebook, Instagram, Youtube, dan Twitter. Dalam penelitian ini, penulis fokus pada analisis data yang ada pada platform twitter karena memiliki karakter utama *text based post* sehingga lebih mudah untuk mengetahui emosi. Selain itu juga terdapat nilai lebih dari sisi kemudahan untuk mengakses API yang diberikan untuk pengembangan penelitian.

Twitter merupakan salah satu sosial media yang memiliki pengaruh cukup baik dalam menaikkan *engagement* pada pelanggan yang dimiliki oleh sebuah *brand*. Salah satu cara yang dilakukan adalah menciptakan format *tweet* yang sesuai dengan identitas *brand* itu sendiri. Bentuk respon yang disediakan oleh twitter adalah retweet, favorite, dan reply. Semakin tinggi respon yang didapatkan, maka akan semakin viral postingan tersebut dan meningkatkan engagement dari konsumen dan calon konsumen yang ada di Twitter. Namun ada 3 jenis sentimen yang terjadi dalam setiap respon, yaitu Positif, Netral, dan Negatif.

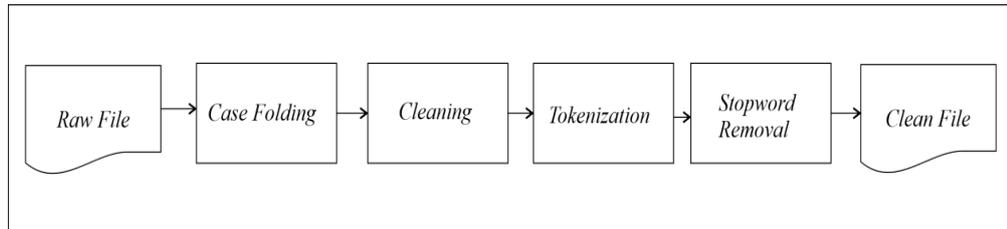
2 METODE PENELITIAN

2.1 Pengumpulan Data

Proses pengumpulan data dilakukan dengan *crawling* data twitter menggunakan R library *Rlang*, akun twitter yang digunakan adalah @gojekindonesia dan @grabID masing-masing mengambil 8.000 *tweet activity*

(*Post*, *Retweet*, dan *Reply*), dilakukan penarikan untuk data pada tanggal 27-31 Agustus 2019.

2.2 Preprocessing Data



Gambar 1 Diagram Alur *Preprocessing* Data Twitter

Dari diagram diatas, dapat dilihat bahwa proses terbagi menjadi 4 tahap, yaitu sebagai berikut.

- Case Folding*, yaitu tahapan untuk menyeragamkan format penulisan dengan menghilangkan kapital dan memasukan huruf latin saja.
- Cleaning*, yaitu proses untuk menghilangkan *noise* pada setiap tweet. Hal yang dibersihkan dalam proses ini adalah duplikasi, URL, *twitter username*, dan tanda baca yang tidak perlu seperti %,/,+,”.
- Tokenization*, adalah proses untuk melakukan pemotongan atau pemisahan string dari kalimat yang terbentuk dalam tweet menjadi token-token kata.
- Stopword removal*, adalah proses untuk menghilangkan *stopword* menggunakan *stoplist*.

Setelah semua tahap dilakukan, maka dihasilkan *file* kumpulan *tweet* yang bersih dan siap untuk dianalisis menggunakan beberapa metode yang telah ditentukan.

2.3 Klasifikasi

Perancangan training untuk melakukan proses klasifikasi dengan menggunakan 3 metode yaitu *Random Forest*, *Support Vector Machine*, dan *Multinomial Naïve Bayes*.

2.3.1. Proses Training

Prosen *training* yang dilakukan pada metode SVM, RF, dan MNB menggunakan langkah sebagai berikut:

- Proses training *Support Vector Machine* memiliki kelebihan dalam melakukan klasifikasi yang akurat dan mampu melakukan generalisasi *high dimensionla features*^[2]. Tahap yang dilakukan adalah memberi label setiap tweet yang telah bersih dengan kelas sen- timennya masing-masing berdasarkan *dictionary* yang digunakan, formula klasifikasi SVM sebagai berikut:

$$f(x) = \sum_{i=1}^n a_i k(x, x_i) + b \tag{1}$$

- Proses training menggunakan *Random Forest* merupakan metode yang populer digunakan karena memiliki performa yang baik dengan

memanfaatkan *tree* yang digabungkan dalam 1 model^[3]. Tahap yang dilakukan adalah melakukan *resampling* menggunakan *bootstrap* dengan jumlah sampel Z^* untuk $b=1$ sampai B , sehingga didapatkan *tree* (T_{b1}^B), setelah terbentuk *tree* tersebut, dilakukan proses klasifikasi dengan persamaan sebagai berikut:

$$C_{rf}^B = \text{Majority Vote}\{C_b(x)\}_1^B \tag{2}$$

- c. Pada proses training menggunakan *Naïve Bayes* tahap yang dilakukan adalah pengkom- binasian peluang setiap fraksi/token yang terbentuk dalam model berikut:

$$\text{Pr}(C) = \log \pi_c + \prod_{w=1}^{|V|} \log(1 + f_w) \log(\text{Pr}(w|c)) \tag{3}$$

2.3.1. Penghitungan Akurasi

Penghitungan akurasi pada proses ini menggunakan *confusion matrix*, dimana proses yang dilakukan adalah membandingkan hasil klasifikasi dari metode yang digunakan dengan hasil klasifikasi yang seharusnya, berikut adalah struktur *confusion matrix*.

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif	FP (<i>False Negative</i>)	TN (<i>True Negative</i>)

Dari hasil yang didapatkan pada *confusion matrix* dapat digunakan untuk menghitung nilai akurasi, presisi, dan recall. Persamaan untuk menghitung nilai-nilai tersebut sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \tag{4}$$

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{FN+TP} \tag{6}$$

Selain menghitung akurasi, juga dilakukan proses penghitungan *Net Sentiment Score* (NSS) yang bertujuan untuk mengetahui kepuasan konsumen dengan menggunakan nilai yang didapat dari proses pengurangan persentase *tweet* negatif dan persentase *tweet* positif^[4]. Rentang nilai NSS adalah -100 sampai 100, semakin tinggi nilai NSS maka semakin puas konsumen terhadap marketing tersebut dan sebaliknya apabila semakin rendah mendekati (-100), maka terjadi indikasi bahwa konsumen kurang puas terhadap marketing yang dilakukan oleh sebuah *brand*.

3 HASIL

Proses pembentukan klasifikasi dengan 3 metode yang telah dirancangan yaitu *random forest*, *support vector machine*, dan *multinomial naïve bayes* telah dilakukan dengan menggunakan data yang telah dibersihkan pada tahap *pre processing* data. Banyaknya data yang digunakan dimasing-masing metode disamakan. Setelah seluruh proses klasifikasi di setiap metode telah dilakukan, kemudian dilakukan penghitungan akurasi disetiap metode dan penghitungan recall dan presisi disetiap kelas positif dan negatif yang didapatkan untuk masing-

masing metode. Hasil dari seluruh proses tersebut dapat dilihat pada table 1 dan tabel 2 berikut

Tabel 1. Nilai Akurasi Klasifikasi

Model	Akurasi
<i>Random Forest</i>	72.35%
<i>Support Vector Machine</i>	76%
<i>Multinomial Naïve Bayes</i>	68.95%

Tabel 2. Evaluasi Model

Kelas	Random Forest		Support Vector Machine		Multinomial Naïve Bayes	
	Precision	Recall	Precision	Recall	Precision	Recall
Positif	0.66	0.52	0.79	0.94	0.45	0.61
Negatif	0.32	0.58	0.72	0.80	0.25	0.78

Berdasarkan hasil dari akurasi dan evaluasi model yang digunakan, maka untuk proses klasifikasi sentimen tweet transportasi daring menggunakan metode support vector machine. Proses selanjutnya adalah melakukan klasifikasi terhadap seluruh data dan membagi hasil tersebut menjadi 2 bagian berdasarkan sumber akun yang mengeluarkan tweet, yaitu @gojekindonesia dan @grabid. Hasil dari klasifikasi sentimen tersebut kemudian digunakan untuk menghitung net sentiment score (nss), dan hasilnya dapat dilihat pada tabel 3.

Tabel 3. Net Sentiment Score

Akun	Persentase Positif	Persentase Negatif	NSS
@gojekindonesia	67	33	34
@GrabID	82	18	64

4 PEMBAHASAN

Hasil yang ditunjukkan pada tabel 1 memperlihatkan bahwa dari ketiga metode klasifikasi yang digunakan untuk menentukan kelas sentimen dalam positif dan negatif, metode *Support Vector Machine* merupakan metode yang paling akurat dengan nilai 76%. Jika dilihat pada evaluasi model yang terdapat pada tabel 2, presisi dan recall untuk kelas negatif lebih rendah dari kelas positif, hal ini menjadi kelemahan dari ketiga metode yang digunakan, oleh karena itu diperlukan penelitian lebih lanjut dalam menentukan metode klasifikasi yang lebih baik. Jika dilihat tujuan penggunaan hasil klasifikasi untuk menghitung NSS yang dihasilkan dari persentase selisih jumlah elemen dari masing-masing kelas, maka akurasi 76% masih kurang baik dan riskan apabila selisih yang dihasilkan bernilai kecil. Namun pada penelitian ini, nilai NSS yang dihasilkan cukup besar yaitu 34 untuk @gojekindonesia dan 64 untuk @GrabID. Dari hasil NSS dapat disimpulkan bahwa kedua perusahaan berhasil memberikan kepuasan kepada konsumen, namun

jika dibandingkan keduanya, maka @GrabID memiliki jumlah konsumen yang puas lebih banyak dibanding- kan dengan @gojekindonesia.

5 KESIMPULAN

1. Transportasi daring merupakan kebutuhan baru bagi masyarakat Indonesia, khususnya pada masyarakat perkotaan.
2. Penggunaan sosial media untuk *creative campaign* merupakan cara yang efektif untuk mendapatkan dan menjaga loyalitas pelanggan.
3. Dari 3 metode klasifikasi yang digunakan dalam penelitian ini, metode klasifikasi *Support Vector Machine* merupakan metode yang paling baik dibandingkan dengan 2 metode lainnya dengan 76% akurasi.
4. Nilai NSS pada akun twitter @GrabID adalah 64 atau lebih tinggi dibandingkan dengan @gojekindonesia yang mendapat nilai 34. Hal ini menunjukkan bahwa jumlah konsumen yang puas dengan produk Grab lebih banyak jika dibandingkan dengan jumlah konsumen yang puas dengan produk gojek yang dikampanyekan melalui sosial media.

DAFTAR PUSTAKA

- [1] Pollay, R.W. & Mittal, B. (1993). Here's the beef: Factors, determinants and segments in consumer criticism of advertising. *Journal of Marketing*, 57(3), 99-114.
- [2] T. Joachims (1997). Text categorization with support vector machines: Learning with many relevant features. Technical Report 23, Universitat Dortmund, LS VIII.
- [3] K. Pearson(1904), "On the theory of contingency and its relation to association and nor- mal correlation," Cambridge University Press, London.
- [4] Ye Q, Zhang Z and Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches *Expert Syst. Appl.* vol36 no3 pp 6527–35.