



Tabel 1.

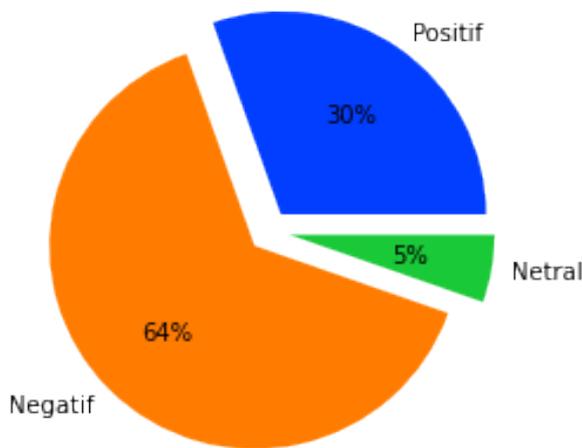
Data aktualdan peramalan IHK menggunakan parameter lain

No.	Komentar
1	Anak" kasian kalo harus antigen.. Tolong di hapuskan dong peraturan untuk anak usia di bawah 12 th harus antigen ???? ribet amat sih peraturannya skrg..
2	saya kira dah cukup double vaksin
3	Sama aj beratt, ,,percuma vaksin udh 2x kl msh hrs keluar duit buat per/antigen!!
4	kenapa transportasi BUS aja yg tdk SWAB,??? APA BEDANYA,???? malah naik kereta anak2 disuruh swab, pemerintah kok tidak searah
dst	...

Tabel 2.

Nilai parameter yang digunakan

Jenis Transportasi	$\alpha(0)$	$R(0)$	$\delta_\alpha$
Kereta Api	0.67	3	0.86
Pesawat	0.2	4	0.94
Kapal	0.79	1	0.59
Bus	0.68	3	0.89



Gambar 3. Komposisi sentimen.

beberapa lembaga dalam penyesuaian kebijakan. Komentar-komentar tersebut dapat diolah untuk mengetahui tanggapan masyarakat terhadap kebijakan yang dibuat dan bisa menjadi bahan evaluasi untuk pembuatan kebijakan selanjutnya. Salah satu cara untuk mengolah data berupa teks adalah *text mining* dimana merupakan topik khusus dari *data mining* yang akan menggali informasi dari data-data berupa teks. Fungsi *text mining* yang akan digunakan dalam penelitian ini adalah pengelompokan teks atau *text clustering* dengan tujuan mengelompokkan komentar masyarakat kedalam beberapa kelompok dan setiap kelompok akan dianalisis. Metode yang digunakan adalah *Self-Organizing Maps* (SOM) yang merupakan salah satu algoritma jaringan syaraf tiruan dan salah satu aplikasinya adalah masalah perjalanan *salesman* [2]. *Self-Organizing Maps* mampu memetakan data berdimensi tinggi ke dalam dimensi rendah sehingga dinilai cocok untuk *text clustering* yang merupakan aplikasi dari penggunaan data berdimensi tinggi [3]. Komentar pada setiap *cluster* dicari sentimennya menggunakan *Indonesian Sentiment Lexicon* dengan mencari polaritas komentar untuk mengetahui positif atau negatifnya tanggapan masyarakat terhadap kebijakan yang ditetapkan.

Pada jurnal ini akan membahas mengenai analisis sentimen menggunakan *Indonesian Sentiment Lexicon* pada komentar

Tabel 3.

Hasil clustering

Jenis Transportasi	Komentar yang Dominan Tiap Cluster	$\delta_\alpha$
Kereta Api	Komentar mengenai genose	SE Kemenhub 27-2021
	Komentar mengenai antigen	SE Kemenhub 89-2021
	Komentar mengenai syarat perjalanan anak-anak	SE Kemenhub 58-2021
	Komentar mengenai vaksin	SE Kemenhub 38-2021
Pesawat	Komentar mengenai genose	SE Kemenhub 57-2021
	Komentar mengenai PCR	SE Kemenhub 88-2021 dan SE Kemenhub 21-2022
	Komentar mengenai peraturan perjalanan secara umum	SE Kemenhub 62-2021
	Komentar mengenai syarat perjalanan anak-anak	SE Kemenhub 45-2021
Kapal	Komentar mengenai vaksin	SE Satgas 17-2021
	Komentar mengenai antigen	SE Satgas 21-2021
	Komentar mengenai syarat perjalanan anak-anak	SE Satgas 22-2021
	Komentar mengenai vaksin	SE Satgas 21-2021
Bus	Komentar mengenai syarat perjalanan anak-anak	SE Satgas 22-2021
	Komentar mengenai peraturan perjalanan secara umum	SE Satgas 21-2021
	Komentar mengenai antigen	SE Kemenhub 43-2021 dan SE Satgas 21-2021

terhadap kebijakan perjalanan domestik yang sudah dikelompokkan menggunakan metode *Self-Organizing Maps*. Komentar pada setiap kelompok dicari sentimen yang dominannya sehingga dapat mengetahui tanggapan masyarakat lebih mengarah pada positif, negatif, atau netral.

## II. TINJAUAN PUSTAKA

### A. Text Mining

*Text mining* adalah proses penggalian informasi dari data-data berupa teks. Tujuan dari *text mining* adalah mendapatkan informasi atau wawasan dari kumpulan data teks. Beberapa penerapan *text mining* yaitu perangkuman teks, pengelompokan teks, klasifikasi teks, dan lain-lain [4]. *Text mining* memerlukan tahapan praproses teks agar bisa diolah lebih lanjut. Tahap pertama dalam praproses teks adalah *case folding*. Pada tahap ini, seluruh karakter teks diubah menjadi huruf kecil serta menghilangkan tanda baca dan karakter selain huruf alfabet. Tahap kedua adalah *tokenizing*, dimana suatu kalimat dipecah menjadi kata-kata. Proses berikutnya yaitu menghapus *stopwords*. *Stopwords* adalah kosakata penghubung atau keterangan dalam sebuah kalimat. Contoh dari *stopwords* dalam bahasa indonesia adalah “yang”, “dari”, “dan”, “atau”, dan lain-lain. Setelah menghilangkan *stopwords*, akan dilakukan proses *stemming*. *Stemming* adalah proses menghilangkan awalan, akhiran, dan sisipan sehingga didapat sebuah kata dasar [5].

### B. Term Frequency-Inverse Document Frequency

*Term Frequency-Inverse Document Frequency* (TF-IDF) adalah proses pembobotan dengan tujuan untuk menemukan bobot setiap kata. Proses *Term Frequency-Inverse Document Frequency* memerlukan dua hal, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* adalah frekuensi suatu kata dalam sebuah dokumen. *Inverse Document Frequency* adalah frekuensi suatu kata pada seluruh dokumen. Berikut adalah rumus untuk menentukan bobot kata dengan TF-IDF [6].

$$tfidf_{ij} = tf_{ij} \times idf_i = tf_{ij} \times \ln\left(\frac{N+1}{df_{i+1}}\right) + 1 \quad (1)$$

dengan:

$tfidf_{ij}$  : bobot kata ke- $i$  terhadap dokumen ke- $j$

$tf_{ij}$  : frekuensi kata ke- $i$  di dokumen ke- $j$

$N$  : jumlah dokumen, dan  $df_i$  frekuensi dokumen yang mengandung kata ke- $i$ .

Selanjutnya TF-IDF dinormalisasikan menggunakan rumus berikut. TF-IDF dapat dinormalisasikan menggunakan rumus berikut [7]:

$$TFIDF_{ij} = \frac{tfidf_{ij}}{\sqrt{\sum_{j=1}^n (tfidf_{ij})^2}} \quad (2)$$

Dengan  $n$  adalah jumlah kata pada seluruh dokumen. Setiap kata pada seluruh dokumen akan dihitung bobotnya. Nilai bobot semakin besar jika suatu kata sering muncul dalam suatu dokumen dan jarang muncul pada dokumen lain [7].

### C. Silhouette Coefficient

Salah satu tugas dalam evaluasi *clustering* adalah penentuan jumlah *cluster*. Metode untuk menentukan jumlah *cluster* salah satunya adalah *Silhouette Coefficient*. Misal sebuah himpunan data  $D$  dengan  $n$  data. Himpunan data  $D$  akan dikelompokkan kedalam  $k$  *cluster*,  $C_1, \dots, C_k$ . Untuk setiap data  $p \in C_i (1 \leq i \leq k)$ , hitung  $a(p)$  sebagai rata-rata jarak antara data  $p$  dengan data lainnya dalam *cluster* yang dimiliki data  $p$ . Demikian pula  $b(p)$  adalah minimum rata-rata jarak antara data  $p$  dengan data lainnya yang tidak satu *cluster* dengan data  $p$ . Jika  $q \in C_i (1 \leq i \leq k)$  dimana  $p \neq q$ ,  $y \in C_j (1 \leq j \leq k)$  dimana  $i \neq j$ ,  $|C_i| = n_i$ ,  $|C_j| = n_j$  maka perumusan  $a(p)$  dan  $b(p)$  adalah sebagai berikut [4].

$$a(p) = \frac{\sum_{x=1}^{n_i} d(p, q_x)}{n_i - 1} \quad (3)$$

$$b(p) = \min \left\{ \frac{\sum_{x=1}^{n_j} d(p, y_x)}{n_j} \right\} \quad (4)$$

Maka rumus dari *Silhouette Coefficient* adalah:

$$s(p) = \frac{b(p) - a(p)}{\max\{a(p), b(p)\}} \quad (5)$$

Rentang nilai *Silhouette Coefficient* adalah antara -1 sampai 1. Nilai  $a(p)$  menunjukkan kekompakan *cluster* yang dimiliki  $p$ . Semakin kecil nilainya, maka semakin kompak *cluster* yang dimiliki atau jarak antar data pada suatu *cluster* semakin dekat. Nilai  $b(p)$  menunjukkan sejauh mana  $p$  dengan *cluster* lainnya. Semakin besar nilainya, maka semakin jauh  $p$  dengan *cluster* lainnya. Maka dari itu, saat *Silhouette Coefficient* semakin mendekati nilai 1, dimana

*cluster* yang dimiliki kompak dan  $p$  jauh dari *cluster* lainnya, maka hasil *clustering* semakin baik. Untuk mencari *Silhouette Coefficient* secara keseluruhan, hitung rata-rata dari *Silhouette Coefficient* seluruh objek pada himpunan data [4].

### D. Self-Organizing Maps

*Self-Organizing Maps* (SOM) adalah salah satu model jaringan syaraf tiruan yang bersifat *unsupervised learning*, dimana tidak membutuhkan label pada proses *training*. *Self-Organizing Maps* biasa digunakan dalam *data mining* yang berkaitan dengan pengelompokan data atau *clustering* [8]. *Self-Organizing Maps* terdiri dari 2 lapisan, yaitu lapisan *input* dan lapisan *output*. Setiap *neuron* pada lapisan *input* dihubungkan ke seluruh *neuron* pada lapisan *output* dengan membawa bobot tertentu. *Neuron* pada lapisan *input* merupakan data masukkan untuk proses *clustering* dan *neuron* pada lapisan *output* merupakan *cluster* yang terbentuk. Selama proses *clustering*, setiap *neuron* pada lapisan *input* akan dihitung jarak kuadratnya dengan bobot. *Neuron* yang memiliki jarak kuadratik terdekat dengan bobot disebut sebagai *neuron* pemenang. *Neuron* pemenang dan *neuron* ketetanggaannya diperbarui bobot-bobotnya, jika ketetanggaannya lebih dari 0. Jika ketetanggaannya 0, maka bobot yang diperbarui hanya *neuron* pemenang. Proses pembaruan bobot dilakukan untuk semua *input*. Arsitektur jaringan *Self-Organizing Maps* secara umum dapat dilihat pada Gambar 1 [2].

Berikut adalah algoritma *Self-Organizing Maps* [2]:

1. Inisialisasi bobot  $w_{ij}$ , atur parameter ketetangaan ( $R$ ), atur parameter *learning rate* ( $\alpha$ ), dan konstanta penurunan *learning rate* ( $\delta_\alpha$ )
2. Untuk setiap vektor input  $x$  yang berjumlah  $n$ , lakukan langkah 3 sampai 5
3. Untuk setiap node output  $j$ , hitung jarak kuadratik *Euclidean* :

$$d(j) = \sum_{i=1}^n (w_{ij} - x_i)^2 \quad (6)$$

4. Tentukan  $j$ , sedemikian sehingga  $d(j)$  minimum
5. Untuk setiap unit  $j$  dengan spesifikasi ketetangaan tertentu dari  $j$ , dan untuk setiap  $i$  maka

$$w_{ij}(\text{baru}) = w_{ij}(\text{lama}) + \alpha [x_i - w_{ij}(\text{lama})] \quad (7)$$

6. Perbarui *learning rate* ( $\alpha$ ) :

$$\alpha(t+1) = \alpha(t) \times \delta_\alpha \quad (8)$$

7. Kurangi  $R$  sebanyak 1 setiap 20 iterasi
8. Jika jumlah iterasi belum mencapai 100, maka kembali ke langkah 2. Jika jumlah iterasi sudah mencapai 100, maka berhenti

### E. Analisis Sentimen

Analisis sentimen adalah bidang penelitian yang bertujuan untuk menganalisis sentimen orang terhadap berbagai topik, peristiwa, individu, masalah, produk, layanan, dan organisasi. Sejak awal 2000-an, analisis sentimen dan penggalian opini sudah mulai diteliti berbagai teknik telah diusulkan untuk menganalisis emosi dan opini dari media sosial [9].

Analisis sentimen yang dilakukan pada penelitian ini menggunakan *Indonesian Sentiment Lexicon*. *Indonesian Sentiment Lexicon* adalah metode analisis sentimen berbasis

kamus data berisi kata-kata sentimen berbahasa Indonesia yang telah disertai bobot pada setiap katanya. *Indonesian Sentiment Lexicon* yang digunakan adalah *lexicon* yang dibuat oleh Fajri Koto dan Gemala Y. Rahmaningtyas dalam penelitiannya yang berjudul “InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs” yang dibangun untuk mengidentifikasi pendapat tertulis dan mengelompokkannya menjadi pendapat positif atau negatif, yang dapat digunakan untuk menganalisis sentimen publik terhadap topik, peristiwa, atau produk tertentu. Dalam *lexicon* tersebut telah terdapat 3609 kata-kata positif dan 6609 kata-kata negatif dengan skor berkisar antara -5 dan +5 [10].

### III. METODE PENELITIAN

#### A. Pengumpulan Data dan Informasi

Pada tahap ini dilakukan pengumpulan data yang diambil dari komentar pada Twitter, Facebook, dan Instagram pada postingan yang membahas syarat perjalanan domestik pada akun resmi PT Angkasa Pura, PT Kereta Api Indonesia, PT Pelayaran Nasional Indonesia, Perusahaan Umum DAMRI, dan Kementerian Perhubungan RI dalam rentang waktu Januari 2021 sampai 8 Maret 2022. Informasi yang dikumpulkan didapat dari situs resmi kementerian perhubungan Republik Indonesia dan situs resmi satuan tugas penanganan COVID-19 yang berisi surat edaran yang membahas mengenai kebijakan perjalanan domestik selama COVID-19.

#### B. Praproses Data

Setelah data terkumpul, akan dilakukan praproses data seperti *case folding* untuk mengubah huruf kapital menjadi huruf kecil dan menghapus karakter selain huruf, *tokenizing* untuk membagi kalimat menjadi kata-kata, normalisasi, menghapus *stopwords*, *stemming* untuk mengambil kata dasar dari suatu kata, dan ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

#### C. Clustering Menggunakan Self-Organizing Maps

Hal pertama yang dilakukan adalah menentukan jumlah *cluster* menggunakan *Silhouette Coefficient*, setelah itu masuk proses *Self-Organizing Maps*. Masukkan untuk *Self-Organizing Maps* adalah matriks TF-IDF berukuran jumlah komentar dikali banyaknya kata dalam seluruh komentar. Keluaran dari *Self-Organizing Maps* adalah *cluster - cluster* berjumlah sesuai hasil dari *Silhouette Coefficient*.

#### D. Analisis Sentimen

Analisis sentimen dilakukan menggunakan *Indonesian Sentiment Lexicon* dengan menghitung polaritas dari setiap komentar.

### IV. ANALISIS HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah komentar pada postingan media sosial Facebook, Instagram, dan Twitter dalam rentang waktu 1 Januari 2021 sampai 8 Maret 2022 yang membahas mengenai kebijakan perjalanan domestik pada akun resmi PT Angkasa Pura, PT Kereta Api Indonesia, PT Pelayaran Nasional Indonesia, Perusahaan Umum DAMRI, dan Kementerian Perhubungan RI. Jumlah

data yang didapat sebanyak 2949 komentar dengan rincian 740 komentar untuk transportasi pesawat, 1636 komentar untuk transportasi kereta api, 285 komentar untuk transportasi kapal, dan 288 komentar untuk transportasi bus. Berikut contoh komentar yang didapat. Beberapa contoh data komentar yang didapatkan dapat dilihat pada Tabel 1.

Fitur yang digunakan dalam penelitian ini adalah komentar dan surat edaran yang dibahas. Komentar terhadap kebijakan transportasi kereta api banyak membahas surat edaran menteri perhubungan nomor SE 89 Tahun 2021. Surat edaran transportasi pesawat yang paling sering dibahas adalah surat edaran menteri perhubungan nomor SE 62 Tahun 2021. Komentar pada kebijakan transportasi kapal paling banyak membahas surat edaran satuan tugas penanganan COVID-19 nomor 17 tahun 2021. Transportasi bus banyak membahas surat edaran satuan tugas penanganan COVID-19 nomor 21 tahun 2021.

#### B. Praproses Data

Data teks yang didapat melalui media sosial bermacam-macam seperti bahasanya sudah baku, menggunakan bahasa gaul serta singkatan, terdapat kesalahan penulisan, dan lain-lain. Agar data teks yang digunakan menjadi sejenis, maka perlu dilakukan tahap praproses teks mulai dari *case folding* sampai tahap *stemming*. Hasil praproses teks pada komentar terhadap kebijakan kereta api dapat dilihat pada Gambar 2.

Pada Gambar 2 (a) terlihat bahwa saat sebelum praproses teks, masih terdapat beberapa kata sambung yang tidak cocok untuk dijadikan fitur. Setelah melalui praproses teks, didapat beberapa kata penting yang cocok dijadikan fitur untuk proses berikutnya seperti yang ditunjukkan pada Gambar 2 (b).

Setelah praproses teks, akan dilakukan ekstraksi fitur menggunakan TF-IDF agar mendapatkan bobot tiap kata untuk dijadikan fitur. Masukkan dari tahap ekstraksi fitur adalah data teks sesudah *stemming*. Keluaran dari tahap ekstraksi fitur menggunakan TF-IDF adalah matriks berukuran banyaknya komentar dikali banyaknya kata dari seluruh komentar. Berikut adalah matriks TF-IDF untuk data kereta yang berukuran  $16363 \times 1452$  dimana 1636 menunjukkan jumlah komentar dan 1452 menunjukkan jumlah kata dari keseluruhan komentar.

$$tfidf_{kereta} = \begin{bmatrix} 0 & \dots & 0.359756054 & \dots & 0 \\ 0 & \dots & 0.419425049 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} \quad (9)$$

#### C. Clustering Menggunakan Self-Organizing Maps

Proses *clustering* diawali dengan menentukan jumlah *cluster* optimal menggunakan metode *Silhouette Coefficient*. Proses penentuan jumlah *cluster* optimal dilakukan dengan menjalankan proses *clustering* dengan beberapa jumlah *cluster* dalam rentang tertentu. Proses *clustering* yang memiliki nilai *Silhouette Coefficient* terbesar akan terpilih menjadi *clustering* yang optimal dengan jumlah *cluster* yang ditentukan. Masukkan dari tahap tersebut adalah matriks TF-IDF yang didapat dari proses sebelumnya. Keluaran dari tahap ini adalah jumlah *cluster* yang optimal untuk proses *clustering*.

Setelah mendapatkan jumlah *cluster* optimal, akan

dilakukan proses *clustering* dengan *Self-Organizing Maps*. Langkah pertama adalah inisialisasi bobot awal secara acak, *learning rate*, ketetanggaan, dan konstanta penurunan *learning rate*. Selanjutnya menghitung jarak kuadrat *Euclidean* antara bobot dengan vektor masukkan (matriks TF-IDF). Setelah mendapatkan semua nilai jarak bobot dengan masukkan, selanjutnya akan dicari *neuron* pemenang dimana memiliki jarak terkecil. Berikutnya adalah memperbarui bobot untuk *neuron* pemenang dan *neuron* tetangganya. Proses mencari *neuron* pemenang dan pembaruan bobot dilakukan untuk setiap data masukkan. Selanjutnya adalah perbarui *learning rate* dan ketetanggaan. *Learning rate* menurun secara linear setiap iterasi. Ketetanggaan juga menurun selama proses *clustering* [2]. Proses dari menghitung jarak kuadrat antara bobot dan data masukkan sampai pembaruan *learning rate* dan ketetanggaan dilakukan sebanyak jumlah iterasi yang ditentukan.

Penelitian ini melakukan 10 kali percobaan menggunakan nilai parameter yang berbeda-beda. Nilai parameter dari setiap jenis transportasi yang menghasilkan *clustering* dengan nilai *Silhouette Coefficient* tertinggi akan digunakan untuk *clustering*. Nilai parameter yang digunakan dapat dilihat pada Tabel 2.

Dengan menggunakan nilai parameter tersebut, penurunan ketetanggaan setiap 20 iterasi, dan jumlah iterasi 100, diperoleh jumlah *cluster* optimal untuk pengelompokan komentar terhadap kebijakan kereta api adalah 4, jumlah *cluster* optimal untuk pengelompokan komentar terhadap kebijakan pesawat adalah 5, jumlah *cluster* optimal untuk pengelompokan komentar terhadap kebijakan kapal adalah 3, dan jumlah *cluster* optimal untuk pengelompokan komentar terhadap kebijakan bus adalah 4. Detail hasil pengelompokan dapat dilihat pada Tabel 3.

Berdasarkan Tabel 3, komentar mengenai syarat perjalanan anak-anak dan vaksin merupakan komentar yang dominan terhadap kebijakan semua jenis transportasi. Genose hanya dominan pada transportasi kereta api dan pesawat. Persyaratan tes genose sudah tidak berlaku sejak surat edaran satuan tugas penanganan COVID-19 nomor 14 tahun 2021 dikeluarkan, yaitu pada 2 Juli 2021. Komentar terhadap kebijakan kapal dan bus banyak diajukan setelah Juli 2021 sehingga hanya sedikit komentar yang membahas mengenai tes genose. PCR hanya dominan pada transportasi pesawat karena terdapat peraturan hanya diperbolehkan tes PCR sebagai bukti hasil negatif COVID 19 yang ditetapkan pada surat edaran menteri perhubungan nomor 57 tahun 2021 pada Juli 2021.

#### D. Analisis Sentimen

Komposisi sentimen dari seluruh komentar pada kebijakan 4 jenis transportasi ditunjukkan pada Gambar 3.

Sentimen negatif merupakan sentimen yang dominan pada komentar terhadap kebijakan perjalanan domestik. Komentar terhadap kebijakan kereta api merupakan komentar dengan sentimen negatif terbanyak, yaitu 1188 komentar atau sekitar 72% dari seluruh jumlah sentimen negatif. Hal yang dominan pada komentar terhadap kebijakan kereta api dan kapal didominasi oleh sentimen negatif. Komentar terhadap kebijakan pesawat didominasi oleh sentimen negatif, kecuali genose yang didominasi oleh sentimen positif dan peraturan perjalanan secara umum yang jumlah sentimen negatif dan

sentimen positifnya seimbang. Komentar terhadap kebijakan bus didominasi oleh sentimen negatif, kecuali pada peraturan perjalanan secara umum yang jumlah sentimen negatif dan sentimen positifnya seimbang.

Komentar terhadap semua surat edaran yang dicantumkan pada Tabel 3 dominan pada sentimen negatif, kecuali pada surat edaran menteri perhubungan nomor SE 38 Tahun 2021 dan surat edaran menteri perhubungan nomor SE 43 Tahun 2021 yang dominan pada sentimen positif. Surat edaran dengan jumlah sentimen negatif terbanyak adalah surat edaran menteri perhubungan nomor SE 89 tahun 2021 yang berisi kebijakan transportasi kereta api dimana komentar terhadap surat edaran tersebut banyak membahas antigen dan syarat perjalanan anak-anak. Surat edaran dengan jumlah sentimen positif terbanyak adalah surat edaran satuan tugas penanganan COVID-19 nomor 17 tahun 2021 yang berisi kebijakan semua jenis. Surat edaran dengan jumlah sentimen netral terbanyak adalah surat edaran menteri perhubungan nomor SE 89 tahun 2021 dan surat edaran menteri perhubungan nomor SE 58 tahun 2021 yang berisi kebijakan transportasi kereta api dimana komentar terhadap kedua surat edaran tersebut banyak membahas antigen, syarat perjalanan anak-anak, dan vaksin.

#### V. KESIMPULAN DAN SARAN

Kesimpulan yang didapatkan dari hasil dan pembahasan adalah sebagai berikut: (1) Metode *Self-Organizing Maps* dapat diterapkan untuk pengelompokan komentar pada media sosial terhadap kebijakan perjalanan domestik dengan langkah-langkahnya adalah menentukan *learning rate*, ketetanggaan, konstanta penurunan *learning rate* serta bobot awal secara acak berbentuk matriks berukuran jumlah fitur  $\times$  jumlah *cluster* yang telah ditentukan menggunakan *Silhouette Coefficient*, menghitung jarak kuadrat antara bobot awal dengan data masukkan yang telah melalui tahap praproses data (*case folding*, *tokenizing*, normalisasi, menghapus *stopwords*, *stemming*, dan ekstraksi fitur TF-IDF) dan mencari indeks jarak kuadrat minimum, memperbarui bobot untuk indeks jarak kuadrat minimum dan tetangganya, memperbarui *learning rate* dan ketetanggaan. Proses tersebut terus berjalan hingga mencapai jumlah iterasi yang telah ditentukan. (2) Pengelompokan komentar terhadap kebijakan kereta api menghasilkan 4 *cluster* dengan komentar yang dominan pada masing-masing *cluster* adalah genose, antigen, syarat perjalanan anak-anak, dan vaksin. Pengelompokan komentar terhadap kebijakan pesawat menghasilkan 5 *cluster* dengan komentar yang dominan pada masing-masing *cluster* adalah genose, PCR, peraturan perjalanan secara umum, syarat perjalanan anak-anak, dan vaksin. Pengelompokan komentar terhadap kebijakan kapal menghasilkan 3 *cluster* dengan komentar yang dominan pada masing-masing *cluster* adalah antigen, syarat perjalanan anak-anak, dan vaksin. Pengelompokan komentar terhadap kebijakan bus menghasilkan 4 *cluster* dengan komentar yang dominan pada masing-masing *cluster* adalah vaksin, syarat perjalanan anak-anak, peraturan perjalanan secara umum, dan antigen. (3) Sentimen komentar terhadap kebijakan perjalanan domestik didominasi oleh sentimen negatif kecuali genose pada transportasi pesawat yang didominasi sentimen positif dan peraturan perjalanan secara umum yang

jumlah sentimen positif dan jumlah sentimen negatif seimbang.

Saran untuk penelitian selanjutnya adalah sebagai berikut:

(1) Dapat menggunakan metode *topic modelling* seperti *Latent Dirichlet Allocation* untuk menemukan topik yang dibahas dari data yang digunakan. (2) Dapat menambahkan tahapan pra-proses data yaitu deteksi pencilan dan seleksi fitur agar data yang akan diproses lebih bersih sehingga bisa diperoleh hasil *clustering* yang lebih baik.

#### DAFTAR PUSTAKA

- [1] A. Nadine and Z. Z. Imtiyaz, "Analisis upaya pemerintah dalam menangani mudik melalui peraturan menteri perhubungan nomor 25 tahun 2020 pada masa covid-19," *Media Juris*, vol. 3, no. 3, pp. 277–298, 2020, doi: 10.20473/mi.v3i3.20674.
- [2] L. Fausett, *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*. New Jersey: Prentice-Hall, 1994.
- [3] Y.-C. Liu, M. Liu, and X.-L. Wang, "Application of Self-Organizing Maps in Text Clustering: A Review," in *Applications of Self-Organizing Maps*, 2012, pp. 205–220. doi: 10.5772/50618.
- [4] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann Publishers, 2012.
- [5] D. S. Indraloka and B. Santosa, "Penerapan text mining untuk melakukan clustering data tweet shopee indonesia," *Jurnal Sains dan Seni ITS*, vol. 6, no. 2, pp. 2337–3520, 2017.
- [6] D. Yogish and R. S. Hegadi, "Variants of term frequency and inverse document frequency of vector space model for effective document ranking in information retrieval," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 7, pp. 414–421, 2019.
- [7] Moch. I. Irawan, R. Wijayanto, M. L. Shahab, N. Hidayat, and A. M. Rukmi, "Implementation of social media mining for decision making in product planning based on topic modeling and sentiment analysis," *J Phys Conf Ser*, vol. 1490, 2020.
- [8] T. A. Milatina, "Implementasi Self-Organizing Fuzzy Maps Pada Incomplete Data Untuk Pengelompokan Gizi Bahan Pangan," Departemen Matematika, Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- [9] S. Rouhani and F. Mozaffari, "Sentiment analysis researches story narrated by topic modeling approach," *Social Sciences & Humanities Open*, vol. 6, no. 1, p. 100309, 2022, doi: <https://doi.org/10.1016/j.ssaho.2022.100309>.
- [10] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *International Conference on Asian Language Processing (IALP)*, 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.