

## Analisis Sentimen Masyarakat terhadap Layanan BPJS Kesehatan dan Faktor-Faktor Pendukung Opini dengan Pemodelan *Natural Language Processing* (NLP)

Dicky Lihardo Girsang \*, Alwi Sidiq, Tahniah Salsabila Elenaputri

IPB University, Jl. Raya Dramaga, Bogor 16680, Indonesia

\*Corresponding author: [dickylihardo@gmail.com](mailto:dickylihardo@gmail.com)



E-ISSN: 2986-4178

### Riwayat Artikel

Dikirim:

29 Maret 2023

Direvisi:

14 Mei 2023

Diterima:

01 Juni 2023

### ABSTRAK

Indonesia secara jelas mengakui dan mewajibkan pemenuhan hak setiap warga negara dalam hal pelayanan kesehatan. BPJS Kesehatan bertugas sebagai penyelenggara jaminan kesehatan yang berjalan sesuai dengan prinsip asuransi sosial dan prinsip ekuitas. Akan tetapi, di samping upaya pemberian pelayanan yang prima untuk setiap pasien, terdapat banyak keluhan masyarakat terhadap layanan BPJS kesehatan, baik itu melalui *call center*, maupun melalui media massa seperti Twitter. Selain sebagai salah satu negara dengan pengguna Twitter terbesar di dunia, adanya pandemi Covid-19 juga serta merta menjadi alasan masyarakat cukup aktif melakukan interaksi lewat sosial media. Oleh karena itu, penelitian ini bertujuan untuk menganalisis polaritas serta kecenderungan opini masyarakat terhadap layanan BPJS Kesehatan secara holistik berdasarkan cuitan publik di Twitter. Data tekstual diperoleh dengan teknik *text mining* didasarkan pada kata kunci yang relevan sebagai gambaran dari layanan BPJS kesehatan yang dirasakan langsung oleh masyarakat. Dari pra-proses data, klasifikasi, pembobotan, hingga menggunakan *probabilistic topic modelling*, yakni dengan metode *Latent Dirichlet Allocation* (LDA), diperoleh informasi bahwa sekitar 6 dari 10 opini masyarakat mayoritas ke arah yang negatif. Adapun hasil ekstraksi menggambarkan tiga topik utama, yakni mahal nya biaya pengobatan, terdapat beberapa biaya yang tidak di-cover oleh BPJS, serta biaya pengobatan spesifik ke dokter gigi.

**Kata Kunci:** BPJS kesehatan, analisis sentimen, *indobert*, *latent dirichlet allocation*.

### ABSTRACT

Indonesia clearly acknowledges and obligates the fulfilment of every citizen's right to healthcare services. BPJS Kesehatan serves as the organizer of health insurance that runs in accordance with the principles of social insurance and equity. However, in addition to efforts to provide excellent service for every patient, there are many complaints from the public regarding BPJS Kesehatan services, both through call centres and mass media such as Twitter. Besides one of the countries with the largest number of Twitter users in the world, the COVID-19 pandemic has also prompted people to interact

*actively through social media. Therefore, this study aims to analyze the polarity and tendencies of public opinion on BPJS Kesehatan services holistically based on public tweets on Twitter. Textual data obtained through text mining techniques based on relevant keywords as a representation of BPJS Kesehatan services perceived directly by the public. From data pre-processing, classification, and weighting, to using probabilistic topic modelling, namely the Latent Dirichlet Allocation (LDA) method, it is found that approximately 6 out of 10 public opinions tend towards negativity. The extraction results describe three main topics: the high cost of treatment, several costs not covered by BPJS, and the cost of specific treatment for dental care.*

**Keywords:** *BPJS kesehatan, sentiment analysis, indobert, latent dirichlet allocation*

## 1. Pendahuluan

Media sosial dapat menjadi sebuah jembatan komunikasi antara suatu instansi dan masyarakat, di mana aspirasi masyarakat berperan penting dalam keberhasilan kebijakan yang dibuat [6]. Salah satu *platform* yang populer bagi masyarakat Indonesia untuk menyampaikan opini mereka adalah media sosial *microblogging*, yaitu Twitter. Berdasarkan data dari We Are Social, pengguna Twitter di Indonesia mencapai angka 18,45 juta pada tahun 2022 atau meningkat sekitar 31,3% dari tahun 2021. Angka tersebut menempatkan Indonesia di peringkat kelima sebagai negara dengan pengguna Twitter terbesar di dunia. Melalui *tweet*, Twitter menjadi wadah bagi masyarakat di Indonesia untuk menyatakan aspirasi mereka terhadap suatu hal, misalnya pelayanan suatu instansi. Pada penelitian ini, akan dilakukan analisis terhadap opini atau persepsi masyarakat terhadap pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Indonesia pada periode dua tahun terakhir, yaitu tahun 2021–2022. Data diambil melalui *tweet* di media sosial Twitter dengan sejumlah kata kunci (*keyword*) yang relevan. Berdasarkan data yang tertera dalam situs resmi BPJS Kesehatan, jumlah peserta Jaminan Kesehatan Nasional (JKN) terus mengalami pertumbuhan hingga mencapai 245,8 juta jiwa per 31 Oktober 2022. Artinya, persentase penduduk Indonesia yang terdaftar sebagai peserta Jaminan Kesehatan Nasional pada tahun 2022 terbilang sangat besar, yaitu mencapai 89%. Oleh karenanya, pertumbuhan peserta tersebut harus diimbangi dengan peningkatan akses layanan kesehatan, baik dari kemudahan akses maupun ketersediaan fasilitas. Publikasi BPJS Kesehatan pada tahun 2019 menyatakan bahwa sistem JKN BPJS Kesehatan menerapkan sistem pelayanan berjenjang. Garda terdepan yang harus dikunjungi peserta BPJS dalam melakukan pengobatan adalah Fasilitas Kesehatan Tingkat Pertama (FKTP) yang meliputi Pusat Kesehatan Masyarakat (puskesmas), klinik pratama, dokter umum, laboratorium, dan jejaring (bidan, laboratorium jejaring, apotek jejaring, dan faskes penunjang yang terkait kerja sama dengan BPJS Kesehatan).

Melalui *tweet*, diperoleh data tekstual yang dapat dianalisis dengan menggunakan analisis sentimen yang merupakan salah satu teknik *text mining* [5]. Analisis Sentimen digunakan untuk menemukan informasi berharga yang dibutuhkan—baik positif, negatif, maupun netral—yang terkandung dalam sebuah data yang tidak terstruktur [4]. Pelabelan dapat dilakukan dengan model yang telah dilatih sebelumnya (*pre-trained model*). *Pre-trained model* dilatih pada set data yang besar untuk mempelajari suatu fungsi sehingga umumnya telah memiliki kualitas yang sangat baik. Dalam analisis ini, digunakan *pre-*

*trained model Bidirectional Encoder Representations from Transformers (BERT)* berbahasa Indonesia atau IndoBERT [9]. Algoritma BERT digunakan untuk mengklasifikasikan informasi dengan teknik analisis sentimen yang dikategorikan menjadi tiga jenis polaritas, yaitu positif, negatif, dan netral [1]. Kurniawan *et al.*, 2021 [7] menyatakan dalam penelitiannya bahwa model IndoBERT memiliki performa terbaik dibandingkan model lain yang diteliti.

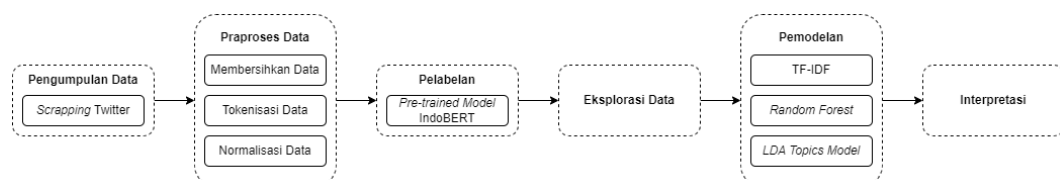
Berbagai opini dari pengguna terhadap suatu produk atau jasa dalam suatu *platform* merupakan bentuk dari kualitas produk atau jasa itu sendiri. Dalam hal ini, *tweet* opini peserta terhadap pelayanan BPJS Kesehatan merupakan bentuk dari kualitas layanan tersebut terhadap pesertanya. Data mentah yang didapatkan dari media sosial memerlukan tahap praproses [3]. Praproses data terdiri atas beberapa tahapan, seperti *cleaning* data, transformasi data, dan reduksi data. Langkah-langkah yang dilakukan dalam tahap praproses data tidak selalu sama, tetapi menyesuaikan dengan kebutuhan data yang ada. Data hasil praproses merupakan data yang sudah bersih dan siap untuk dianalisis.

Data tekstual yang telah melewati tahap praproses diubah menjadi numerik untuk menjalankan proses klasifikasi. Pengubahan data menjadi numerik dilakukan dengan metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*). Perhitungan bobot terdiri atas dua konsep, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF merupakan frekuensi munculnya suatu kata dalam suatu dokumen, sedangkan IDF merupakan logaritma perbandingan banyaknya dokumen dengan frekuensi munculnya suatu kata dalam dokumen [13]. Dari hasil pembobotan tersebut, dilakukan penentuan *feature importance* menggunakan *random forest*, lalu dilanjutkan dengan pemodelan topik.

Pemodelan topik dilakukan menggunakan model probabilitas *Latent Dirichlet Allocation* (LDA), yang umumnya digunakan untuk menganalisis dokumen dalam ukuran besar [2]. Keluaran LDA adalah distribusi topik untuk setiap dokumen yang selanjutnya dapat digunakan sebagai masukan untuk klusterisasi. Melalui penerapan beberapa algoritma tersebut, dilakukan analisis terhadap keluaran sehingga diperoleh informasi mengenai sentimen masyarakat Indonesia terhadap pelayanan BPJS Kesehatan pada tahun 2021–2022 di media sosial Twitter.

## 2. Metodologi Penelitian

Berikut adalah tahapan yang dilakukan untuk mendapatkan informasi mengenai sentimen masyarakat Indonesia terhadap pelayanan BPJS Kesehatan pada tahun 2021–2022 di media sosial Twitter dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

### 2.1. Pengumpulan Data

*Dataset* dibangun dengan teknik *web scraping* dari Twitter dengan sejumlah kata kunci (*keyword*) yang relevan terhadap topik, yaitu pelayanan BPJS, biaya BPJS, dokter BPJS, BPJS lama, dan BPJS membantu. Kata kunci tersebut diambil dari periode Januari

2021 hingga November 2022 dan didapatkan 56.388 baris data dengan lima kolom sebagai peubah. Contoh hasil *scraping* dapat dilihat pada Tabel 1.

**Tabel 1.** Contoh hasil *scraping*

<i>Date Created</i>	<i>Username</i>	<i>Tweets</i>	<i>Like Count</i>	<i>Location</i>
2022-10-22 02:40:06+00:00	oconizm	@babecabiita Kalo masalah lama mah wajar lah, antrian BPJS nya juga panjang	0	Bandung, West Java, Indonesia
2022-11-04 06:39:24+00:00	Budenovi	Saya senang dg BPJS. Pasang 2 ring ditanggung semua. Pelayanan bagus di rumah sakit @aricanti dan @SiloamBali . Kalau antri sy rasa wajar. <a href="https://t.co/rlmLyTfVrD">https://t.co/rlmLyTfVrD</a>	1	Ubud, Bali, Indonesia
2022-11-07 04:13:11+00:00	BPJSKesehatanRI	@LonyyyLove Salam Sahabat. Mohon maaf atas ketidaknyamanannya. Dalam hal penyediaan pelayanan obat merupakan kewenangan pemberi pelayanan kesehatan (faskes) yang diawasi oleh BPOM sesuai Formularium Nasional dan e-katalog yang ditetapkan oleh pemerintah. Terima kasih. :) -lia	0	Indonesia
2022-11-03 12:13:39+00:00	SahiL2AW	@liktono @gibran_tweet Sy rasa masyarakat berhak mendapatkan pelayanan terbaik, kalau sudah mampu di gratiskan lebih baik gratis. Kalau perlu kesehatan untuk rakyat juga gratis, tidak perlu bayar bpjs. Memang idealnya buat rakyat itu demikian, jangan semua di bebaskan ke rakyat	3	Indonesia

## 2.2. Praproses Data

### 2.2.1. Data cleaning

- *Remove duplicate*, yaitu penghapusan terhadap data yang berduplikasi.
- *Remove tweet from BPJS Kesehatan RI*, yaitu penghapusan keyword berupa *username @BPJS Kesehatan RI* yang tidak mengandung opini publik.
- *Remove number*, yaitu penghapusan terhadap seluruh angka pada data.
- *Remove mention and special character*, yaitu penghapusan *mention*, *link*, dan karakter khusus.
- *Case folding*, yaitu penggantian seluruh abjad pada data menjadi huruf kecil.

- *Remove punctuation*, yaitu penghapusan terhadap tanda baca.
- *Remove single char*, yaitu penghapusan terhadap huruf tunggal.
- *Remove multiple whitespace*, yaitu penghapusan karakter kosong berupa beberapa spasi.

Setelah dilakukan proses *data cleaning* dengan urutan seperti di atas, didapatkan hasil sebagai berikut.

**Tabel 2.** Contoh hasil *data cleaning*

<i>Tweet</i>	@babecabiita Beda RS beda pelayanan sih. Dari 3 RS yg biasa sy datengin, yg 1 emg kayak kirang ramah bpjs, ruang tunggu di luar, lebih lama. 2 RS lain alhamdulillah ga ada komplain.
<i>Remove number</i>	@babecabiita Beda RS beda pelayanan sih. Dari RS yg biasa sy datengin, yg emg kayak kirang ramah bpjs, ruang tunggu di luar, lebih lama. RS lain alhamdulillah ga ada komplain.
<i>Remove mention and special character</i>	Beda RS beda pelayanan sih. Dari RS yg biasa sy datengin, yg emg kayak kirang ramah bpjs, ruang tunggu di luar, lebih lama. RS lain alhamdulillah ga ada komplain.
<i>Case folding</i>	beda rs beda pelayanan sih. dari rs yg biasa sy datengin, yg emg kayak kirang ramah bpjs, ruang tunggu di luar, lebih lama. rs lain alhamdulillah ga ada komplain.
<i>Remove punctuation</i>	beda rs beda pelayanan sih dari rs yg biasa sy datengin yg emg kayak kirang ramah bpjs ruang tunggu di luar lebih lama rs lain alhamdulillah ga ada komplain
<i>Remove single char</i>	beda rs beda pelayanan sih dari rs yg biasa sy datengin yg emg kayak kirang ramah bpjs ruang tunggu di luar lebih lama rs lain alhamdulillah ga ada komplain
<i>Remove multiple whitespace</i>	beda rs beda pelayanan sih dari rs yg biasa sy datengin yg emg kayak kirang ramah bpjs ruang tunggu di luar lebih lama rs lain alhamdulillah ga ada komplain

### 2.2.2. Tokenisasi

Tokenisasi adalah suatu proses pemecahan sebuah kalimat menjadi kepingan-kepingan kata sesuai dengan kata penyusunnya. Potongan kata tersebut dikenal sebagai token atau *term*.

**Tabel 3.** Contoh hasil tokenisasi

<i>Tweets</i>	beda rs beda pelayanan sih dari rs yg biasa sy datengin yg emg kayak kirang ramah bpjs ruang tunggu di luar lebih lama rs lain alhamdulillah ga ada komplain
Tokenisasi	'beda', 'rs', 'beda', 'pelayanan', 'sih', 'dari', 'rs', 'yg', 'biasa', 'sy', 'datengin', 'yg', 'emg', 'kayak', 'kirang', 'ramah', 'bpjs', 'ruang', 'tunggu', 'di', 'luar', 'lebih', 'lama', 'rs', 'lain', 'alhamdulillah', 'ga', 'ada', 'komplain'

### 2.2.3. Normalisasi

Proses normalisasi terdiri atas dua tahapan, yaitu mengganti *slang words* menjadi kata baku dan menghapus *stopwords*. Untuk mengganti *slang words* menjadi kata baku, kamus (*slang dictionary*) yang digunakan merupakan hasil penelitian Ibrohim dan Budi (2018) tentang pendeteksian Bahasa Indonesia yang bersifat menghina atau *abusive*, di mana kata-kata tersebut biasanya menyinggung kondisi, binatang, bagian tubuh, makhluk astral, dan sebagainya. Contoh Bahasa Indonesia yang bersifat menghina atau *abusive* tersebut adalah *gila, bego, banci, anjing*, dan lainnya. Selanjutnya, dilakukan penghapusan *stopwords* yang merupakan kumpulan kata yang banyak ditemukan dalam Bahasa Indonesia, tetapi tidak memiliki makna dalam kalimat, seperti *adalah, aku, dan beberapa* [11]. Penghapusan *stopwords* dilakukan dengan tujuan menghilangkan kata-kata yang memiliki informasi rendah dalam sebuah teks sehingga analisis dapat difokuskan terhadap kata-kata yang mengandung informasi penting.

**Tabel 4.** Contoh hasil normalisasi

<i>Tweets</i>	beda rs beda pelayanan sih dari rs yg biasa sy datengin yg emg kayak kirang ramah bpjs ruang tunggu di luar lebih lama rs lain alhamdulillah ga ada komplain
<i>Replace slang word dan stopwords removal</i>	beda rumah sakit beda pelayanan rumah sakit datengin kayak kurang ramah bpjs ruang tunggu lama rumah sakit alhamdulillah tidak komplain

### 2.3. Pelabelan

IndoBERT disebut sebagai modifikasi dari BERT Base yang memiliki 12 *hidden layers* dan total lebih dari 220M kata yang dilatih. Model *Natural Language Processing* (NLP) yang diimplementasikan pada *library Transformer*, IndoBERT, menyertakan dua mekanisme terpisah, yakni *encoder* untuk membaca input teks serta *decoder* untuk menghasilkan nilai prediksi. Fitur di dalamnya bekerja dengan membaca seluruh urutan kata dengan sekaligus, bukan sekuensial dari kiri ke kanan. Algoritma pelatihan model IndoBERT terdiri atas tahapan *fine tuning* pada tugas tertentu yang spesifik seperti klasifikasi sentimen hingga didapatkan *output* yang optimal dalam tugas tersebut. Setelah didapatkan *output* terbaik dari *fine tuning*, perlu dilakukan evaluasi terhadap model tersebut dengan menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Implikasinya, proses pembelajaran model lebih kompleks untuk membaca kontekstual antarkata yang dapat meningkatkan nilai akurasi dari hasil klasifikasi setiap label positif, negatif, maupun netral.

### 2.4. Pemodelan

Dalam penelitian sebelumnya [10], diperoleh hasil uji bahwa akurasi rata-rata dari hasil klasifikasi memiliki nilai yang lebih baik saat menggunakan *feature extraction* TF-IDF dibandingkan dengan klasifikasi *baseline* yang hanya menggunakan *Random Forest*. Metode TF-IDF digunakan untuk perhitungan bobot tiap kata (*term*) yang umum digunakan untuk mendapatkan informasi yang relevan mengenai topik yang dianalisis (*information retrieval*). Metode tersebut bekerja dengan menentukan frekuensi kemunculan dari sebuah kata yang terdapat pada sebuah dokumen dan kebalikan frekuensi dokumen yang di dalamnya terdapat kata tersebut. TF-IDF dihitung dengan menggunakan rumus pada Persamaan 1.

$$W_{TF*IDF(t_i,d_j)} = tf_{t_i,d_j} \times (1 + \log \frac{D}{df_{(t_i)}}) \quad (1)$$

di mana  $W_{TF*IDF(t_i,d_j)}$  adalah bobot *term i* dalam dokumen *j*,  $tf_{t_i,d_j}$  adalah jumlah *term i* dalam dokumen *j*,  $D$  adalah jumlah seluruh dokumen, serta  $df_{(t_i)}$  adalah jumlah seluruh dokumen yang mengandung *term i*.

Selanjutnya setiap sentimen diubah ke dalam bentuk *dummy*, yaitu “positif”: 1 dan “negatif”: 0. Menggunakan peubah respon dan peubah prediktor, dilakukan pemodelan menggunakan algoritma *random forest* untuk menentukan *feature importance*. Peubah respon dan peubah prediktor didapatkan dari hasil pembobotan menggunakan metode TF-IDF. Peubah respon merupakan status sentimen setiap dokumen dan peubah prediktor berupa bobot setiap kata dalam dokumen. Adapun keluarannya berupa kata-kata yang menjadi penciri klasifikasi suatu kata apakah bernilai positif atau negatif. Hasil yang diperoleh yakni berupa skor yang menunjukkan *Gini importance* kata tersebut. *Feature importance* dari kata yang ada akan menghasilkan nilai *Gini importance* (MDI), yaitu rata-rata total penurunan variabel dalam *node impurity* yang diboboti dengan proporsi sampel yang mencapai *node* pada setiap pohon dalam *random forest*.

Setelah analisis menggunakan *feature importance*, dilakukan pemodelan topik dengan metode LDA yang memiliki kemampuan mendeteksi jumlah topik yang ada pada dokumen menjadi kelompok klusterisasi kata [12]. Dalam melakukan analisis tersebut digunakan *packages Gensim* dan *pyLDAvis* untuk melihat keterkaitan antarkata yang dihasilkan oleh sentimen negatif yang didapat dari hasil *pre-trained model* IndoBERT.

### 3. Hasil dan Pembahasan

#### 3.1. Pelabelan *Tweets* dengan IndoBERT

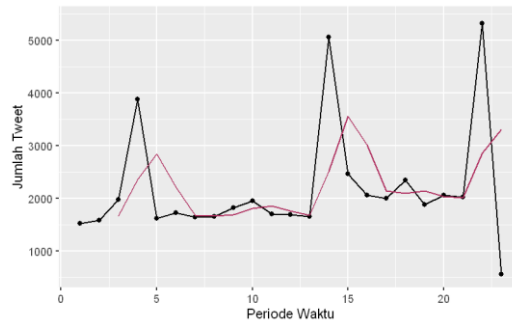
Dengan menggunakan *pre-trained model* IndoBERT, didapatkan label setiap *tweet* yang dapat dilihat pada Tabel 5.

**Tabel 5.** Contoh hasil pelabelan

<i>Text</i>	<i>Status</i>	<i>Score</i>
buka sender periksa minus mata dokter pakai bpjs	<i>neutral</i>	0.99848491
pelayanan rumah sakit jelek menyalahkan bpjs heran banget	<i>negative</i>	0.998391032
senang bpjs pasang ring ditanggung pelayanan bagus rumah sakit antre wajar	<i>positive</i>	0.9573636651

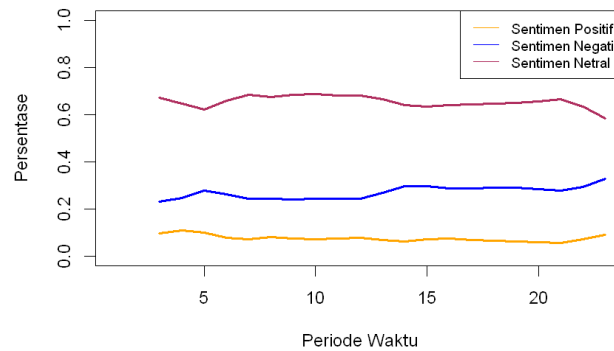
#### 3.2. Eksplorasi Data

Setelah melewati beberapa tahapan dalam praproses data pada hasil *scraping* Twitter berdampak pada berkurangnya jumlah daftar kata dari 56.388 *tweets* menjadi 50.150 *tweets* bersih. Berkurangnya data tersebut diakibatkan dari penghapusan pada duplikasi data dan *tweets* resmi dari *username* @BPJSKesehatanRI yang tidak mewakili opini publik.



**Gambar 2.** Plot sebaran jumlah *tweets* per bulan periode Januari 2021 - November 2022

Secara observatif, *tweet* masyarakat terhadap pelayanan BPJS Kesehatan pada periode waktu yang digunakan paling banyak ditemukan pada bulan Oktober 2022, diikuti bulan Februari 2022 dengan jumlah sekitar 5000 *tweets*. Selain itu, dilakukan pemulusan terhadap jumlah *tweet* per bulan periode Januari 2021 - November 2022 dengan metode *Double Moving Average* (DMA), yakni menghitung rata-rata bergerak sebanyak dua kali, dengan tujuan mengurangi keteracakan data deret berkala. Selain itu, fluktuasi yang cukup tinggi juga menjadi alasan dilakukannya pemulusan pada data.



**Gambar 3.** Plot perbandingan persentase sentimen per bulan periode Januari 2021–November 2022 dengan pemulusan

Berdasarkan data yang digunakan, persentase tertinggi dari polarisasi opini masyarakat Indonesia terhadap pelayanan BPJS Kesehatan periode Januari 2021–November 2022 adalah sentimen netral yang mencapai sekitar 60% setiap bulannya, diikuti oleh sentimen negatif dan sentimen positif di sekitar 30% dan 10% secara berurutan. Sebagian besar masyarakat pengguna Twitter menuliskan opini netral yang mengindikasikan pelayanan BPJS Kesehatan memiliki kelebihan dan kekurangan masing-masing dari setiap aspek. Namun, terlepas dari opini netral, hal tersebut menandakan bahwa masyarakat Indonesia lebih banyak yang menuliskan opini negatif tentang pelayanan BPJS Kesehatan dibandingkan opini positif di media sosial Twitter. Dengan demikian, perlu dilihat hal apa saja yang banyak dikeluhkan oleh masyarakat pengguna BPJS Kesehatan sebagai inovasi terbentuknya pelayanan yang lebih baik kedepannya.

### 3.3. Pemodelan

Metode TF-IDF menghasilkan *output* berupa matriks bobot *term* pada tiap dokumen yang dapat dilihat pada Tabel 6. Frekuensi kata mewakili setiap teks dari data dalam bentuk matriks dengan dimensi barisnya adalah status sentimen masing-masing *tweets* dan kolomnya adalah jumlah kata yang berbeda di dalam korpus dokumen.

**Tabel 6.** *Sparse Matrix* TF-IDF Vector dengan dimensi (13146 × 22317)



Status	adil	administrasi	...	aduh	...
0	0	0	...	0.2371	...
1	0	0.1896	...	0	...
...	...	...	...	...	...
0	0	0	...	0	...

Pengukuran kepentingan setiap kata dalam korpus terhadap klasifikasi sentimen positif dan negatif didasarkan oleh nilai *feature importance*. Setelah diurutkan, didapatkan sepuluh peubah dengan nilai MDG paling besar ditunjukkan pada tabel berikut.

**Tabel 7.** Nilai *Feature Importance* Metode *Random Forest*

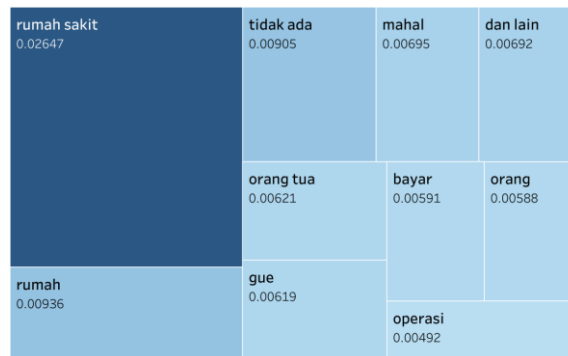
No	Peubah	MDG	No	Peubah	MDG
1	membantu	0.26488	6	banget	0.016546
2	bagus	0.019594	7	semoga	0.014098
3	alhamdulillah	0.019079	8	lama	0.013472
4	pelayanan	0.019027	9	tidak	0.012967
5	bpjs	0.017535	10	terbaik	0.011443

Berdasarkan *feature importance* di atas, kata *membantu* merupakan peubah yang sangat penting terhadap klasifikasi sentimen positif atau negatif yang sudah dilakukan dengan bantuan *pre-trained* IndoBERT. Nilai *Mean Decrease in Gini* (MDG) diperoleh dari hasil perhitungan *information gain* masing-masing variabel dengan menggunakan nilai *entropy* atau *gini* dari pohon yang dihasilkan oleh model. Pada tabel tersebut, terdapat dua diantaranya *feature* untuk sentimen negatif, yakni *lama* dan *tidak*.

Akan tetapi, pendekatan dengan nilai *feature importance* cenderung bias akibat tendensi pada fitur kontinu atau variabel kategorik yang terlalu banyak jumlahnya. Secara kontras, indikasi bias tersebut dapat dilihat dari ketidaksinkronan hasil eksplorasi dengan tabel 7. Dari hasil *scraping*, diperoleh jumlah opini negatif lebih banyak secara kuantitas. Terdapat kemungkinan bahwa fitur-fitur yang tidak secara langsung terkait dengan sentimen negatif dapat berinteraksi dengan fitur-fitur yang berkaitan dengan sentimen negatif. Interaksi ini dapat mempengaruhi prediksi sentimen secara keseluruhan. Namun, ketika melihat *feature importance* secara terpisah, interaksi antara fitur-fitur tersebut mungkin tidak terlihat secara langsung. Oleh karena itu, diperlukan analisis lebih lanjut untuk menemukan pola topik pada data secara holistik.

Kemudian dilakukan analisis lanjutan pada data hasil pelabelan bantuan IndoBERT, spesifik hanya sentimen negatif dengan akurasi di atas 0.95. Pemilihan opini negatif untuk analisis dengan metode LDA ini didasarkan pada pentingnya evaluasi dari program pelayanan tersebut sebagai bentuk pembaharuan di masa mendatang. Penentuan banyaknya model topik dilakukan dengan melihat nilai *coherence score*, semakin tinggi maka semakin baik. Diperoleh bahwa jumlah topik dengan koherensi skor tertinggi adalah sebanyak empat. Akan tetapi, secara garis besar, topik keempat sudah diwakili oleh ketiga

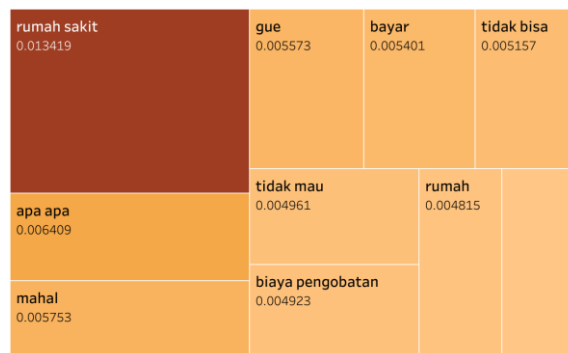
topik lainnya. Oleh karena itu, *num topics* yang dipilih adalah sebesar tiga. Visualisasi *treemap* di bawah ini merupakan *top* sepuluh hasil klusterisasi masing-masing topik.



**Gambar 4.** Treemap chart topik 1

Menyimpulkan *term* yang terdapat pada Gambar 4, topik tersebut membicarakan tentang mahal biaya di rumah sakit apabila pengobatan, termasuk tindakan operasi, tidak menggunakan BPJS Kesehatan. Walaupun mahal, masyarakat terkadang lebih memilih untuk menjadi pasien umum daripada pasien BPJS karena kurang baiknya pelayanan faskes terhadap pasien BPJS. Padahal, kemudahan prosedur yang saat ini diberikan yakni peserta BPJS dapat menggunakan Kartu Indonesia Sehat (KIS) atau hanya menunjukkan KTP pada saat pendaftaran berobat.

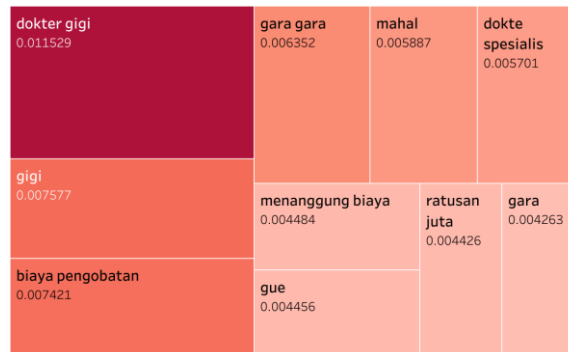
Akan tetapi, pelayanan kesehatan pada peserta BPJS dari masa ke masa tidak berhenti menuai polemik. Seringkali pasien yang belum pulih dipulangkan dengan alasan kuota BPJS yang sudah penuh, juga tidak bisa mendapatkan perawatan karena kamar kelas 1 penuh sehingga disarankan untuk naik kelas dengan status pasien umum. Melihat kejadian pada tahun 2021 ke bawah, buruknya pelayanan mungkin terjadi karena keuangan BPJS Kesehatan yang mengalami defisit sehingga memiliki tunggakan klaim yang telah jatuh tempo kepada rumah sakit. Dengan menyelesaikan masalah tersebut, kualitas pelayanan rumah sakit terhadap pasien BPJS dapat mumpuni.



**Gambar 5.** Treemap chart topik 2

Hasil yang hampir serupa ditunjukkan oleh topik ke-2. Topik tersebut membahas tentang biaya pengobatan yang tidak di-*cover* oleh BPJS di beberapa rumah sakit, termasuk pengobatan pertama (ringan). Hukumnya, rumah sakit yang bekerjasama dengan BPJS Kesehatan tidak dapat menolak pasien BPJS sesuai dengan Pasal 29 Ayat (2) UU No. 44 tahun 2009 tentang rumah sakit. Namun, keinginan masyarakat untuk bisa sembuh cepat seringkali membuat masyarakat lebih memilih untuk berobat tanpa layanan BPJS. Hal tersebut mengindikasikan bahwa masyarakat yang memiliki gejala ringan dan tidak perlu mengeluarkan biaya terlalu besar lebih memilih untuk berobat dengan biaya pribadi

daripada menggunakan BPJS Kesehatan yang pelayanannya terkesan memakan waktu. Menurut Pratiwi *et al.* (2020) [8], pemberian penyuluhan mengenai swamedikasi pada pasien BPJS dapat menumbuhkan kesadaran masyarakat untuk lebih memperhatikan kesehatan. Swamedikasi diartikan sebagai suatu upaya untuk mengobati gejala penyakit tertentu tanpa melakukan konsultasi terlebih dahulu dengan dokter.



Gambar 6. Treemap chart topik 3

Topik yang lebih spesifik ditunjukkan pada klasifikasi topik terakhir, yakni terkait biaya untuk berobat ke dokter gigi. Menurut Peraturan BPJS Kesehatan No. 1 Tahun 2014 Pasal 52 Ayat 1, cakupan layanan perawatan gigi dan mulut yang ditanggung meliputi pramedikasi, konsultasi, pencabutan gigi, ekstraksi, penambalan, dan *scaling* gigi pada periode tertentu. Layanan tersebut telah mencakup perawatan yang umumnya dibutuhkan oleh masyarakat Indonesia. Akan tetapi, masih terdapat keluhan mengenai perawatan gigi dengan menggunakan BPJS Kesehatan akibat kuota yang tidak memadai dari faskes terkait. Dikutip dari siaran pers BPJS Kesehatan pada tahun 2021, BPJS Kesehatan berkomitmen untuk memastikan peserta JKN-KIS mendapatkan layanan kesehatan yang memuaskan sehingga wajib dilakukan seleksi terhadap faskes yang akan bekerja sama. Oleh karena itu, diperlukan peninjauan ulang, baik dari segi fasilitas maupun sumber daya, terhadap faskes yang bekerja sama dengan BPJS Kesehatan.

#### 4. Kesimpulan

Berdasarkan hasil pelabelan menggunakan *pre-trained model* IndoBERT, didapatkan proporsi sentimen positif sekitar 10% setiap bulannya, kemudian sentimen negatif dan sentimen netral masing-masing di sekitar 30% dan 60%. Terlepas dari sentimen netral, masyarakat dominan memiliki sentimen negatif tentang pelayanan BPJS Kesehatan di media sosial Twitter periode tahun 2021–2022. Sentimen negatif tersebut memiliki pengaruh penting terhadap evaluasi pelayanan BPJS Kesehatan sebagai pembaharuan di masa mendatang. Dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk sentimen negatif, diperoleh tiga topik yang meliputi mahal biaya pengobatan di rumah sakit, biaya pengobatan yang tidak ter-cover oleh BPJS di beberapa rumah sakit, serta biaya pengobatan spesifik ke dokter gigi.

#### 5. Daftar Pustaka

- [1] M. F. Abdussalam, D. Richasdy, and M. A. Bijaksana, “BERT implementation on news sentiment analysis and analysis benefits on branding”, *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2064–2073, 2022, doi: 10.30865/mib.v6i4.4579.
- [2] B. W. Arianto and G. Anuraga, “Pemodelan topik pengguna Twitter mengenai aplikasi Ruangguru”, *Jurnal ILMU DASAR*, vol. 21, no. 2, pp. 149–154, 2020, doi: 10.19184/jid.v21i2.17112.

- [3] A. Aziz, Sistem pengklasifikasian entitas pada pesan twitter menggunakan ekspresi regular dan naive bayes, Bogor: Institut Pertanian Bogor, 2013.
- [4] A. P. Giovani, Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis sentimen aplikasi ruang guru di twitter menggunakan algoritma klasifikasi”, *Jurnal TEKNOINFO*, vol. 14, no. 2, pp.116–124, 2020, doi: 10.33365/jti.v14i2.679.
- [5] M. O. Ibrohim and I. Budi, “A dataset and preliminaries study for abusive language detection in indonesian social media”, *Procedia Computer Science*, vol. 135, pp. 222–229, 2018, doi: 10.1016/j.procs.2018.08.169.
- [6] A. N. Izzati, A. Pratama, I. G. A. A. M. Aristamy, N. F. Najwa, and M. A. Rakhmawati, “Kategorisasi jenis interaksi pemerintah dan masyarakat serta popularitas media sosial pemerintah daerah”, *Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 68–70, 2018, doi: 10.21609/jsi.v14i1.567.
- [7] M. Kurniawan, K. Kusriani, and M. R. Arief, “Part of speech tagging pada teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo”, *Jurnal Eksplorasi Informatika*, vol. 11, no. 1, pp. 29–37, 2021, doi: 10.30864/eksplorasi.v11i1.506.
- [8] Y. Pratiwi, A. Rahmawaty, and R. Islamiyati, “Peranan apoteker dalam pemberian swamedikasi pada pasien BPJS”, *Jurnal Pengabdian Kesehatan*, vol. 3, no. 1, pp. 65–72, 2020, doi: 10.31596/jpk.v3i1.69.
- [9] H. K. Putra, M. A. Bijaksana, and A. Romadhony, “Deteksi penggunaan kalimat abusive pada teks Bahasa Indonesia menggunakan metode IndoBERT”, *Jurnal Tugas Akhir Fakultas Informatika*, vol. 8, no. 2, pp. 3028–3038, 2021, doi: 10.33005/jifosi.v2i1.260.
- [10] R. G. Ramli and Y. Sibaroni, “Klasifikasi topik twitter menggunakan metode random forest dan fitur ekspansi word2vec”, *e-Proceeding of Engineering*, vol. 9, no. 1, pp. 79–91.
- [11] A. Santosa, I. Purnamasari, and R. Mayasari, “Pengaruh *stopwords removal* dan *stemming* terhadap performa klasifikasi teks komentar kebijakan *new normal* menggunakan algoritma LSTM”, *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 1, pp. 81–93, 2022.
- [12] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, “ Analisis topik tagar COVID Indonesia pada Instagram menggunakan *latent dirichlet allocation*”, *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.14421/jiska.2022.7.1.
- [13] M. M. Sya'bani and R. Umilasari, “Penerapan metode *cosine similarity* dan pembobotan TF/IDF pada sistem klasifikasi sinopsis buku di Perpustakaan Kejaksaan Negeri Jember”, *Jurnal Sistem & Teknologi Informasi Indonesia (JUSTINDO)*, vol. 3, no. 1, pp. 31–42, 2018