BIG DATA CHALLENGE (BDC)
SATRIA DATA 2022
Universitas Islam Indonesia

Analisis Sentimen Terhadap Kinerja BPJS pada Data Twitter dan Instagram Berdasarkan Model Bert

Analisis Sentimen Sosial Media

BDC_SD202200004

1. Pendahuluan

Latar Belakang

Negara yang maju adalah negara yang masyarakatnya sehat. Masyarakat yang sehat akan bekerja dengan baik sehingga roda perekonomian dapat terus berputar dan negara menjadi maju. Sebaliknya, jika suatu negara memiliki banyak masyarakat yang tidak sehat maka perekonomian suatu negara dapat terhambat karena angka kematian angkatan kerja yang meningkat dan masyarakat yang tidak dapat bekerja dengan baik. Selain itu, berdasarkan dasar negara Indonesia yaitu Pancasila, pada sila ke 2 dan sila ke 5 yang berbunyi ”Kemanusian yang adil dan beradab” dan “Keadilan sosial bagi seluruh rakyat Indonesia,” mempertegas kebutuhan jaminan sosial adalah wajib untuk setiap masyarakat di Indonesia. Oleh karena itu, jaminan sosial di bidang kesehatan yaitu asuransi kesehatan, merupakan sesuatu yang sangat penting bagi kemajuan negara dan keberlangsungan hidup rakyat.

Berdasarkan Undang-Undang nomor 24 Tahun 2011, BPJS Kesehatan (Badan Penyelenggara Jaminan Sosial Kesehatan) merupakan badan yang bertanggung jawab untuk menyelenggarakan program jaminan sosial di bidang kesehatan. Maka dari itu, keberhasilan penyelenggaraan program jaminan sosial di bidang kesehatan sangat bergantung dengan kinerja BPJS Kesehatan. Untuk mengetahui kinerja BPJS Kesehatan, perlu dikumpulkan data opini masyarakat Indonesia mengenai kinerja BPJS. Data tersebut berasal dari *platform social media.* Pada penelitian ini, kami akan mengumpulkan data dari *platform twitter* dan *instagram* karena berdasarkan survei yang dilakukan oleh *hootsuite* yang dilakukan pada Februari 2022, *platform instagram* dan *twitter* menempati urutan ke dua dan ke enam dari social media yang paling sering digunakan di Indonesia. Alasan lebih khususnya memilih *instagram* dan *twitter*, untuk *instagram* jumlah *followers* akun BPJS Kesehatan pada instagram memiliki angka paling besar dibanding *social media* yaitu 1 Juta sedangkan untuk *twitter* merupakan social media yang paling sering digunakan untuk memberi opini tentang isu – isu di suatu negara. Selanjutnya akan dilakukan *sentiment analysis* pada data opini BPJS Kesehatan dari *twitter* dan *instagram* untuk memberi gambaran bagaimana kepuasan pelanggan terhadap kinerja BPJS Kesehatan. Dalam melakukan *sentiment analysis*, tentunya

diperlukan model matematika yang cukup canggih sehingga dapat memberi interpretasi terhadap opini pelanggan yang berbahasa Indonesia secara akurat, salah satu model yang dapat menjalankan tugas ini adalah model BERT. Beberapa kelebihan model BERT adalah sudah dilakukan pretrained pada bahasa Indonesia dan model BERT memiliki akurasi yang bagus karena sering dilakukan update pada model BERT, selain itu model BERT juga sudah terkenal memiliki track record yang bagus dalam bidang NLP. Oleh karena itu, kelompok kami memutuskan menggunakan model BERT dalam melakukan *sentiment analysis* Model BERT pretrained dalam bahasa Indonesia.

Setelah dilakukan *sentiment analysis* pada data dari *Twitter* dan *Instagram* maka dapat diketahui tingkat kepuasan masyarakat Indonesia terhadap pelayanan BPJS di Indonesia lalu dapat dilakukan analisis secara lebih detil untuk mengetahui yang perlu diperbaiki dan dipertahankan dari pelayanan BPJS Kesehatan. Selain data dari *social media*, untuk menilai kinerja BPJS Kesehatan, akan digunakan juga data yang berasal dari panitia *satria data* yang merupakan data biaya tagihan BPJS Kesehatan yang akan berguna untuk memberi gambaran persebaran pengguna BPJS Kesehatan dan persebaran biaya tagihan BPJS Kesehatan berdasarkan kategori tertentu. Dengan demikian, hasil analisis diharapkan akan berguna untuk BPJS Kesehatan agar dapat berkembang lebih baik lagi dalam memberi pelayanan jaminan sosial kesehatan masyarakat Indonesia.

Rumusan Masalah

1. Bagaimana opini masyarakat tentang pelayanan BPJS Kesehatan pada periode 2021-2022?
2. Apa yang menyebabkan kinerja BPJS Kesehatan dinilai buruk oleh pelanggan BPJS?
3. Bagaimana persebaran peserta BPJS di Indonesia?
4. Bagaimana persebaran biaya tagihan BPJS berdasarkan provinsi, jenis fakes, segmen pekerja, dan kepemilikan faskes?

Tujuan Penelitian

1. Menjelaskan kinerja BPJS Kesehatan pada periode 2021-2022.
2. Menjelaskan hal yang menyebabkan kinerja BPJS Kesehatan dinilai buruk oleh pelanggan BPJS.
3. Menjelaskan persebaran peserta BPJS di Indonesia.
4. Menjelaskan persebaran biaya tagihan BPJS berdasarkan provinsi, jenis fakes, segmen pekerja, dan kepemilikan faskes.

2. Metodologi

2.1. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers atau BERT adalah salah satu State of the Art model pembelajaran mesin untuk bidang NLP yang dikembangkan oleh peneliti Google AI Language pada tahun 2018 (Devlin, 2018). Model ini menggunakan transformer, yakni model yang menggunakan mekanisme attention yang dapat mempelajari relasi kontekstual antar kata dalam teks. Transformer ini menggunakan arsitektur encoder-decoder yang mengandung dua model RNN yang bekerja secara simultan dalam menerima input barisan kata dan mengeluarkan output berupa inferensi dari barisan kata/kalima tersebut. Berbeda dengan model-model

terdahulu, model BERT melakukan pelatihan *bidirectional* transformer, sehingga dapat memperoleh konteks yang lebih baik dari teks yang diberikan. Ilustrasi dari kerja model ini dapat dilihat pada diagram di bawah.

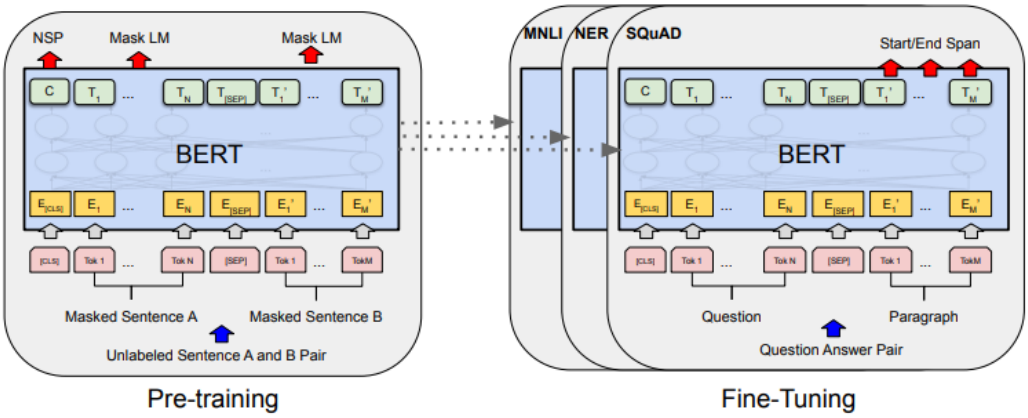


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

Gambar 1. *Alur Kerja Model BERT*

Model ini melakukan pre-training dan fine-tuning secara simultan. Terdapat 2 teknik utama yang digunakan oleh model ini yakni:

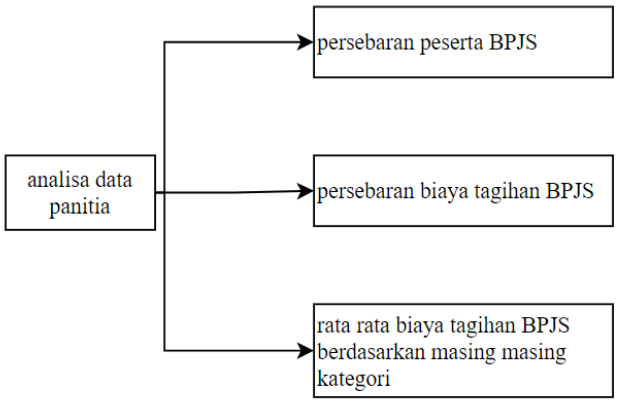
1. *Masked Language Modelling* (Mask LM),
   Diberikan kata tersembunyi dari suatu teks, model BERT akan memprediksi kata tersebut berdasarkan konteks kata lainnya pada teks tersebut,
2. *Next Sentence Prediction* (NSP),
   Model BERT menerima sepasang kalimat secara random dari data teks tersebut sebagai input lalu memprediksi apabila kalimat kedua pada pasangan tersebut merupakan kalimat yang berkesinambung dengan kalimat pertama.

BERT mencetak prestasi sebagai model terbaik dalam berbagai subbidang NLP seperti *question answering*, *Natural Language Inference*, dan terutama analisis sentiment. Riset menunjukkan bahwa model BERT yang diaplikasikan pada dataset *sentiment treebank* menghasilkan skor General Language Understanding Evaluation (GLUE) sebesar 94.9 (Batra, 2021).

Pada penelitian ini digunakan model BERT base untuk memprediksi sentimen dari setiap pesan di social media Instagram dan Twitter.

2.2. Alur Kerja

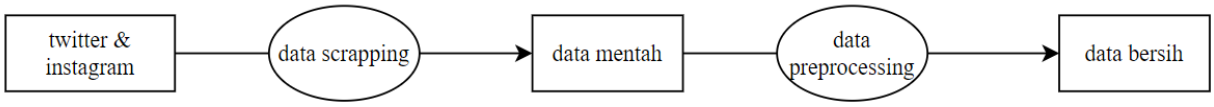
2.2.1. Data Panitia



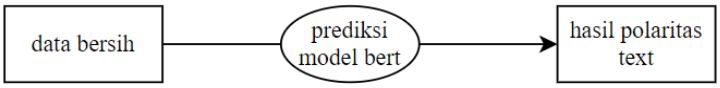
***Gambar 2**. Alur Kerja untuk Data Panitia*

Pada data yang diberikan panitia akan langsung dilakukan analisis karena data bersih sehingga tidak perlu dilakukan proses preprocessing lagi secara mendalam. analisis dibagi menjadi tiga bagian yaitu persebaran peserta BPJS, persebaran biaya tagihan BPJS, dan rerata biaya tagihan BPJS berdasarkan masing masing kategori. Alasan kami menetapkan menjadi tiga bagian seperti pada *gambar 2* adalah untuk mengetahui apakah pelayanan BPJS Kesehatan di Indonesia sudah merata dan apakah terdapat kategori masyarakat atau faskes tertentu yang memiliki biaya tagihan BPJS Kesehatan yang tidak wajar.

2.2.2. Data Twitter dan Instagram



***Gambar 3**. Alur Kerja Bagian 1 pada Data Twitter dan Instagram*



***Gambar 4.** Alur Kerja Bagian 1 pada Data Twitter dan Instagram*

Berdasarkan *gambar 3* dapat dilihat bahwa data mentah yang merupakan opini masyarakat tentang bpjs kesehatan, berasal dari twitter dan instagram, data tersebut akan dibersihkan melalui berbagai proses yang detailnya dapat dilihat pada bagian *3.2* dan *3.3*. Selanjutnya data bersih akan melewati proses *gambar 4* yaitu memprediksi sentimen dari data bersih, pada makalah ini kami menetapkan tiga jenis sentimen yaitu positif, negatif, dan netral. Setelah didapat hasil sentimen berdasarkan model Bert, akan dilakukan berbagai Analisis yang nantinya akan berguna untuk BPJS Kesehatan untuk mengetahui kinerja BPJS Kesehatan dan hal-hal yang menyebabkan BPJS Kesehatan dinilai negatif oleh asyarakat Indonesia

3. Pembahasan

3.1. Data Panitia

3.1.1. Deskripsi Data

Data yang akan dianalisis adalah file yang bernama *2019202004_nonkapitasi.dta*. Data tersebut terdiri dari 166607 baris dan 21 kolom. Data ini tidak memiliki *missing value* dan data tidak duplikat. Fitur-fitur yang ada di data tersebut, yaitu

#	Column	Non-Null Count	Dtype
0	Nomor Peserta	166607 non-null	int32
1	Nomor keluarga	166607 non-null	int32
2	Bobot	166607 non-null	float32
3	ID Kunjungan	166607 non-null	object
4	Tanggal kunjungan	166607 non-null	datetime64[ns]
5	Tanggal tindakan	166607 non-null	datetime64[ns]
6	Tanggal pulang	166607 non-null	datetime64[ns]
7	Provinsi faskes	166607 non-null	category
8	Kode Kab/Kota faskes	166607 non-null	category
9	Kepemilikan faskes	166607 non-null	category
10	Jenis faskes	166607 non-null	category
11	Tipe faskes	166607 non-null	category
12	Tingkat layanan	166607 non-null	category
13	Segmen peserta	166607 non-null	category
14	Kode Nama Diagnosis ICD 10	166607 non-null	category
15	Kode Diagnosis ICD 10	166607 non-null	object
16	Kode Diagnosis	166607 non-null	object
17	Nama Diagnosis	166607 non-null	object
18	Nama Tindakan	166607 non-null	category
19	Biaya tagih	166607 non-null	int32
20	Biaya verifikasi	166607 non-null	int32

3.1.2. Data Kategori

Beberapa fitur yang bertipe *category* memiliki proporsi nilai berupa persentase sebagai berikut.

Jenis faskes

LABORATORIUM	54.207206
PUSKESMAS	25.455713
KLINIK PRATAMA	11.799624
DOKTER UMUM	6.004550
JEJARING	2.532907

Segmen peserta

PPU	28.365555
PBPU	25.091383
PBI APBN	24.243279
BUKAN PEKERJA	15.681214
PBI APBD	6.618569

Tipe faskes

LABORATORIUM	54.207206
RAWAT INAP	14.865522
NON RAWAT INAP	10.590191
KLINIK NON RAWAT INAP	7.856813
DOKTER PRAKTER PERORANGAN	6.004550
KLINIK RAWAT INAP	3.894794
PPK LAIN-LAIN	2.532907
RS KELAS D PRATAMA	0.048017

Kepemilikan faskes

SWASTA	70.616481
PEMERINTAH KABUPATEN/KOTA	26.172370
BUMN	2.687162
TNI AD	0.223280
PEMERINTAH PROVINSI	0.132647
TNI AL	0.083430
POLRI	0.078028
TNI AU	0.006602

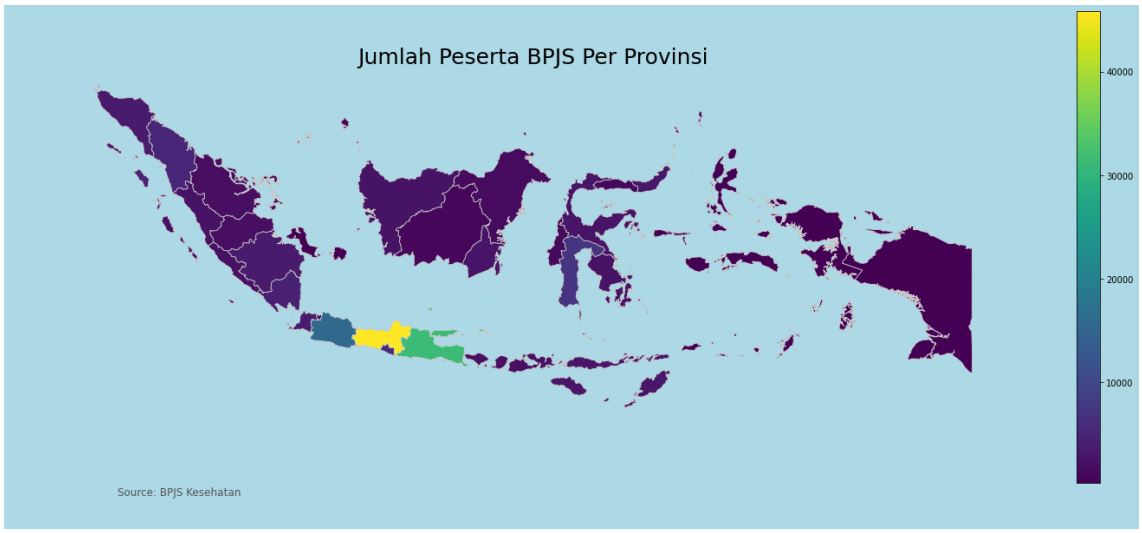
Tingkat layanan

PROMOTIF	63.403098
RJTP	22.976225
RITP	13.620676

Berdasarkan proporsi nilai antara fitur “Tipe faskes” dan “Jenis faskes” tersebut, diperoleh kesimpulan sebagai berikut

- Kategori “LABORATORIUM” pada fitur “Jenis faskes” adalah kategori “LABORATORIUM” pada fitur “Tipe faskes.”
- Kategori “DOKTER UMUM” pada fitur “Jenis faskes” adalah kategori “DOKTER PRAKTEK PERORANGAN” pada fitur “Tipe faskes”.
- Kategori “JEJARING” pada fitur “Jenis faskes” adalah kategori “PPK LAIN-LAIN” pada fitur “Tipe faskes”.
- Kategori “PUSKESMAS” pada fitur “Jenis faskes” adalah kategori “RAWAT INAP” dan “NON RAWAT INAP” pada fitur “Tipe faskes”.
- Kategori “KLINIK PRATAMA” pada fitur “Jenis faskes” adalah kategori “KLINIK NON RAWAT INAP”, “KLINIK RAWAT INAP”, dan “RS KELAS D PRATAMA”.

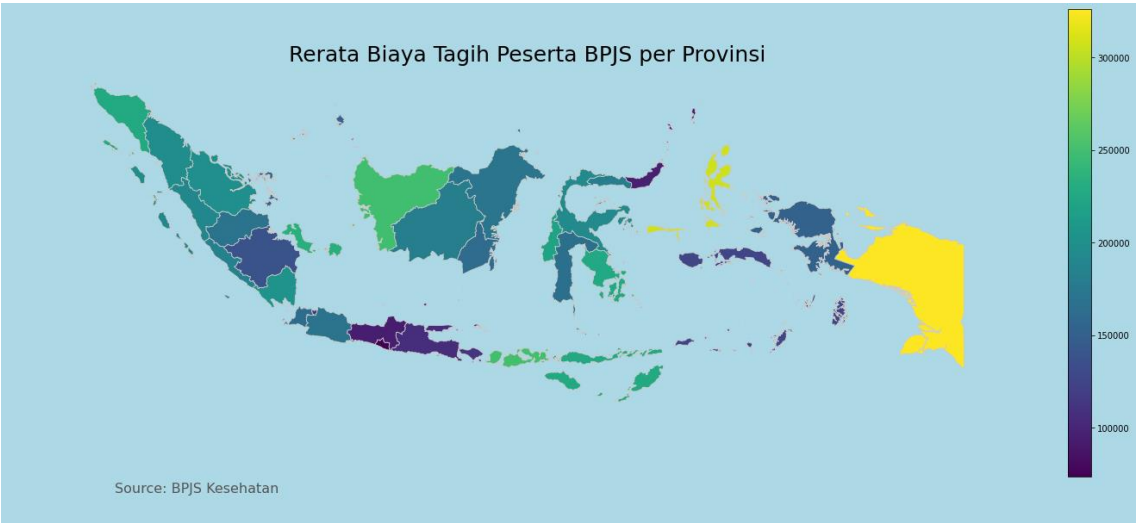
3.1.3. Persebaran Jumlah Peserta BPJS Kesehatan



Gambar 5. *Heatmap Jumlah Peserta per Provinsi*

Berdasarkan *gambar 5*, peserta BPJS terpusat di pulau Jawa. Provinsi dengan jumlah peserta terbanyak adalah Jawa Tengah, sejumlah 45899 peserta. Kemudian, Jawa Timur sejumlah 31410 peserta dan Jawa Barat sejumlah 15531 peserta. BPJS Kesehatan harus mulai berfokus pada pulau selain Jawa agar pelayanannya lebih merata.

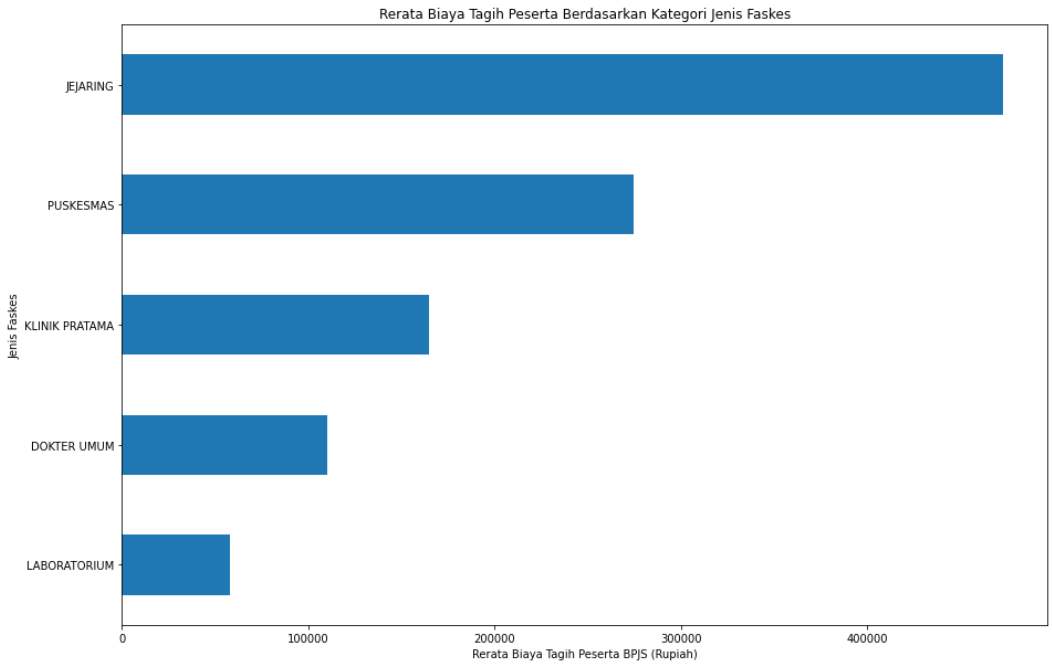
3.1.4. Persebaran Biaya Tagihan BPJS Kesehatan



Gambar 6. *Heatmap Persebaran Biaya Tagihan BPJS per Provinsi*

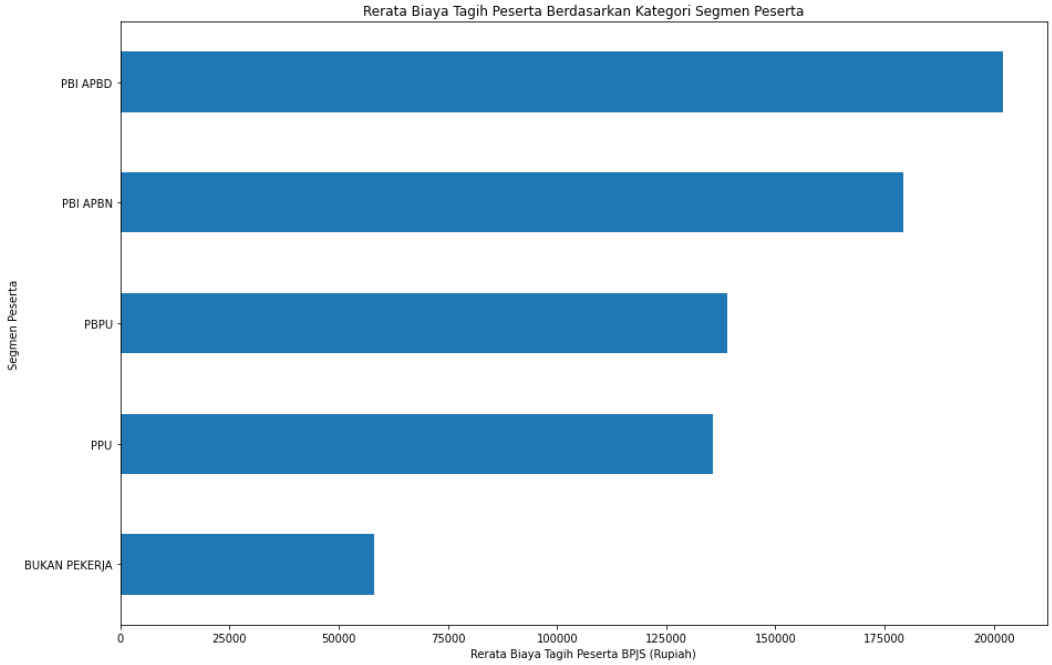
Berdasarkan *gambar 6*, rerata biaya tagih tertinggi peserta BPJS terdapat di provinsi Papua. Kemudian, rerata biaya tagih tertinggi kedua adalah provinsi Maluku Utara. Dari gambar tersebut, pulau Jawa relatif memiliki rerata biaya tagih peserta BPJS yang rendah.

3.1.5. Rerata Biaya Tagihan BPJS Berdasarkan Kategori



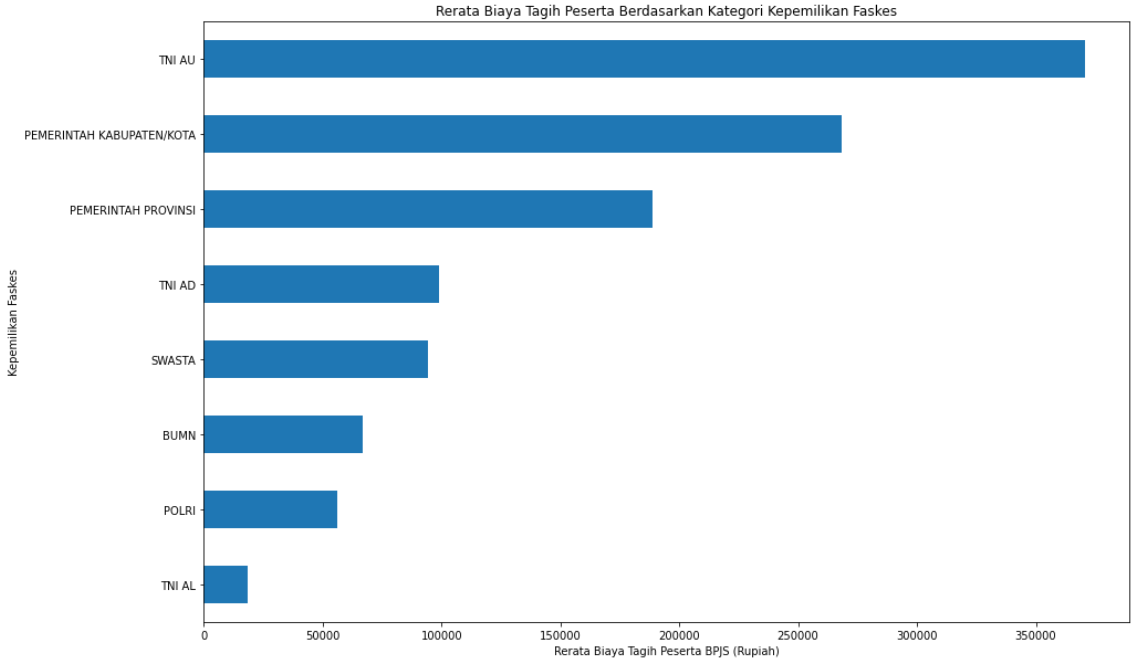
Gambar 7. *Rerata Biaya Tagih Peserta BPJS Berdasarkan Kategori Jenis Faskes*

Berdasarkan *gambar 7*, rerata biaya tagih peserta BPJS tertinggi adalah jenis faskes jejaring.



Gambar 8 *Rerata Biaya Tagih Peserta BPJS Berdasarkan Kategori Segmen Peserta*

Berdasarkan *gambar 8*, rerata biaya tagih peserta BPJS tertinggi adalah segmen peserta PBI APBD.



Gambar 9. *Rerata Biaya Tagih Peserta BPJS Berdasarkan Kategori Kepemilikan Faskes*

Berdasarkan *gambar 9*, rerata biaya tagih peserta BPJS tertinggi adalah kepemilikan faskes oleh TNI AU.

3.2. Data Twitter

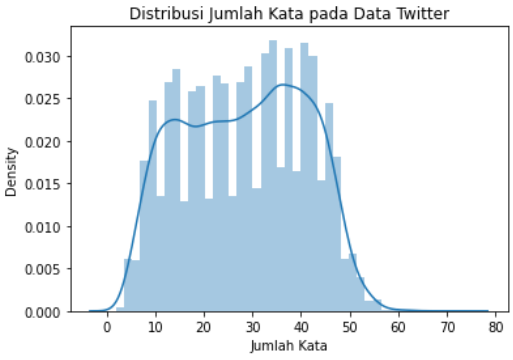
3.2.1. Deskripsi Data

Data untuk social media twitter diperoleh dengan *web scraping* menggunakan *library* snscrape pada python dan mencari post yang mengandung kata ‘bpjs’ dari tahun 2021 hingga tahun 2022. Dengan algoritma tersebut diperoleh 16018 post *tweet* yang mengandung kata BPJS di dalamnya. Adapun, tampilan 5 baris teratas dari data tersebut adalah sebagai berikut,

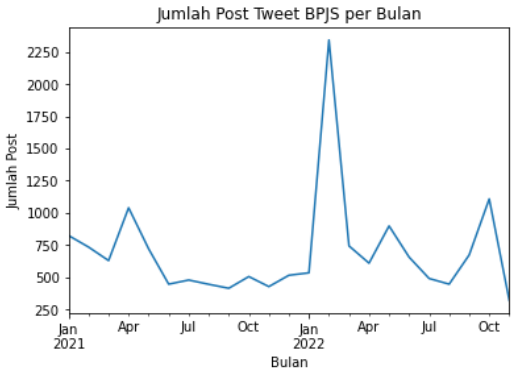
	Date	Tweet
0	2022-11-09 23:54:10+00:00	@kakaknyaEXO @aeribase Pake BPJS boleh gak sih...
1	2022-11-09 14:01:35+00:00	@JafYaya2150 @53NN0_ @SoffiAgus @cagubnyinyir2...
2	2022-11-09 12:46:53+00:00	@reusites @tanyakanrl + bpjs itu membengkak gr...
3	2022-11-09 09:28:38+00:00	Kampanye program BPJS “Gotong Royong, Semua Te...
4	2022-11-09 09:28:36+00:00	[Utas] #BPJSBidikPemuda\n\nPengamat kebijakan ...
...	...	...
16013	2021-01-01 01:31:33+00:00	Hutang Pemerintah kepada Muhammadiyah melalui ...
16014	2021-01-01 00:49:14+00:00	IURAN BPJS RESMI NAIK DITAHUN 2021. \n\nFYI : ...
16015	2021-01-01 00:42:32+00:00	Iuran tarif program jaminan kesehatan nasional...
16016	2021-01-01 00:26:52+00:00	Pak @jokowi yang kami hormati, kami harap agar...
16017	2021-01-01 00:16:32+00:00	Siap-siap, iuran BPJS kesehatan hingga bea met...

16018 rows × 2 columns

3.2.2. Hasil Analisis Data Twitter



***Gambar 10**. Distribusi Jumlah Kata pada Data Twitter*



***Gambar 11.** Jumlah Post tweet BPJS tiap bulan*

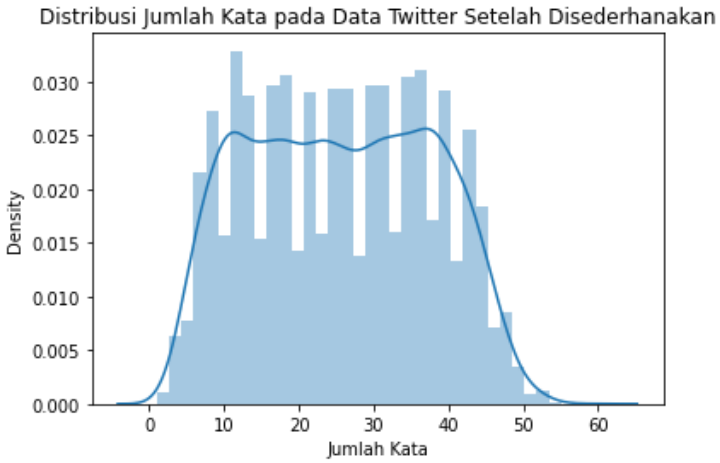
3.2.3. Data Cleaning

Untuk mempercepat komputasi model BERT dalam memprediksi sentimen dari setiap teks tweet, maka teks tweet perlu disederhanakan dengan mengikuti prosedur sebagai berikut:

i. Menghilangkan referensi ke orang tertentu, sebagai contoh @kakaknyaEXO, @reusites)

ii. Menghilangkan link *tweet* yang ada pada akhir post *tweet*
iii. Menghilangkan karakter '#' dan '\n'

Setelah mengikuti prosedur di atas, diperoleh distribusi jumlah kata sebagai berikut,



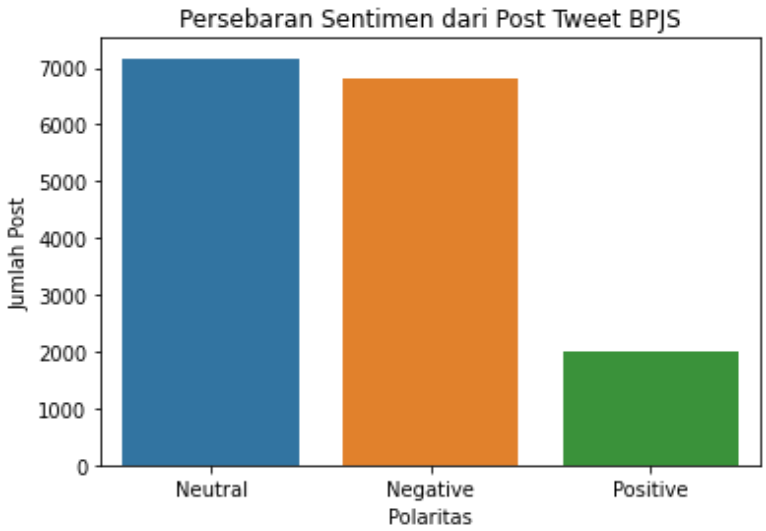
Gambar 12. *Distribusi Jumlah Kata pada data twiter setelah dibersihkan*

Jumlah kata pada post tweet kini berkisar antara 10 - 40 kata

3.2.4. Sentiment Analysis Berdasarkan BERT base

Teks tweet yang telah disederhanakan lalu digunakan sebagai input ke model BERT base. Model BERT base akan mengeluarkan prediksi sentimen/sentimen berdasarkan input yang diberikan. Kami menggunakan *library* HuggingFace dalam python untuk menerapkan model BERT base ini pada setiap post tweet di data tersebut.

Dengan menggunakan model BERT base diperoleh persebaran sentimen pada data post tweet sebagai berikut,



Gambar 13. *Persebaran Sentiment Pada Post Tweet BPJS*

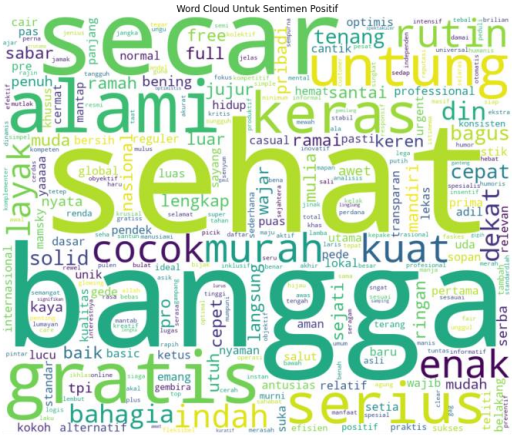
Dapat dilihat bahwa mayoritas post tweet memiliki sentimen netral diikuti oleh sentimen negatif dan paling rendah adalah post tweet dengan sentimen positif.

3.2.5. Analisis Kata Adjektif

Untuk mendapatkan kata-kata adjektif yang menggambarkan setiap sentiment, teks tweet disederhanakan lebih lanjut dengan prosedur sebagai berikut:

i. Menghilangkan karakter non-Alfanumerik,
ii. Menghilangkan kata-kata yang sering muncul (*Stop words*)
iii. Stemming kata dengan menggunakan Lemmatizer Sastrawi
iv. Memperoleh kata-kata adjektif dengan menggunakan POS-Tagging

Setelah menerapkan prosedur di atas, diperoleh WordCloud untuk sentimen positif dan negatif sebagai berikut,



Gambar 14. *Wordcloud Kata Adjektif Positif untuk Tweet Positif*

Gambar 15. Wordcloud Kata Adjektif Negatif untuk Tweet Negatif

Dari *word cloud* sentimen positif**,** dapat dilihat bahwa kata ‘sehat’, ‘bangga’, ‘cocok’, ‘murah’, ‘gratis’, dan ‘untung’ banyak disebut di post tweet bertema BPJS. Sementara, pada *word cloud* tweet negatif dapat dilihat bahwa kata ‘miskin’, ‘tolol’, ‘salah’, dan ‘mahal’ banyak disebut di post tweet bertema BPJS. Selain itu, terdapat kata-kata yang bertolak belakang seperti kata ‘gratis’ pada Word Cloud sentimen positif dan kata ‘mahal’ pada *word cloud* sentimen negatif.

Dapat diduga bahwa beberapa orang berpendapat biaya pengobatan dengan BPJS cukup mahal serta pelayanan BPJS masih belum baik dari segi prosedur. Lalu, fakta bahwa banyak tweet yang dilabeli negatif menandakan bahwa pelayanan BPJS masih dinilai kurang baik oleh sebagian besar masyarakat.

3.3. Instagram

3.3.1. Deskripsi Data

Data didapat berasal dari akun resmi instagram BPJS Kesehatan dari tahun 2021-2022, dengan features yang discrapping adalah

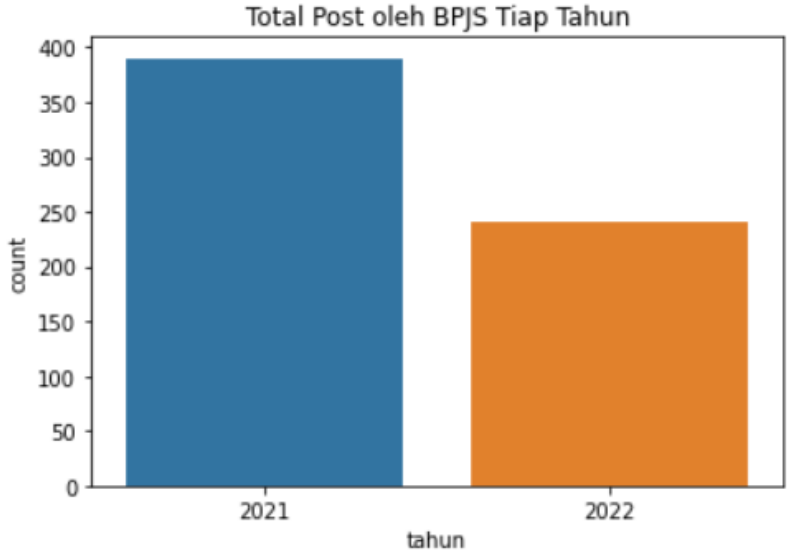
- Konteks: Judul Post
- Waktu_Post: Tanggal dan jam saat suatu post di upload
- Komentar: Komentar masyarakat Indonesia pada post bersangkutan

Total jumlah data yang di scraping pada instagram adalah 35454

	Konteks	Waktu_Post	Komentar
0	Sahabat, ayo segera bayarkan iuranmu sebelum t...	2022-11-07 11:56:53	Sahabat, ayo segera bayarkan iuranmu sebelum t...
1	Sahabat, ayo segera bayarkan iuranmu sebelum t...	2022-11-07 11:56:53	Jelaskan apa itu RITL banyak masyarakat gak ta...
2	Sahabat, ayo segera bayarkan iuranmu sebelum t...	2022-11-07 11:56:53	Maaf kartu bpjs saya kan jatah dari kelurahan....
3	Sahabat, ayo segera bayarkan iuranmu sebelum t...	2022-11-07 11:56:53	Min mgkn perlu dibenahi pelayanannya dlu. Saya...
4	Sahabat, ayo segera bayarkan iuranmu sebelum t...	2022-11-07 11:56:53	Min mau tanya bisa tidak mengulang program reh...

3.3.2. Jumlah Post Instagram Tiap Tahun

Total post akun BPJS Kesehatan tahun 2021 dan 2022 adalah 631 dengan 391 post pada 2021 dan 241 post pada 2022.



Gambar 16. *Jumlah Post Instagram Pada Tahun 2021-2022*

3.3.3. Data Cleaning

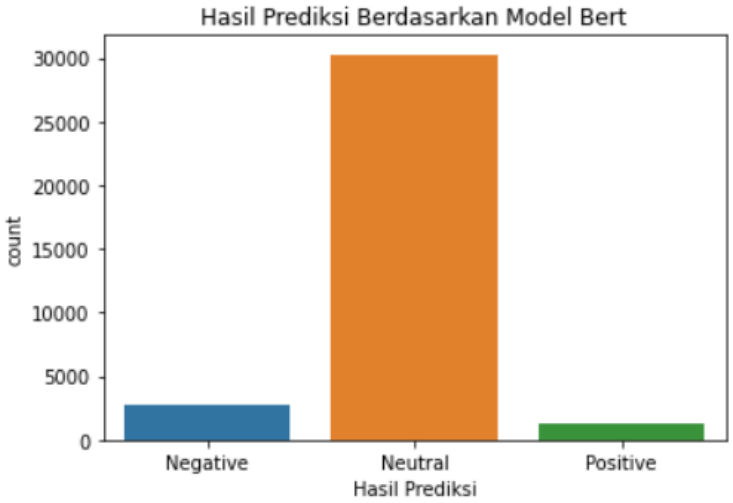
Selanjutnya pada features komentar dilakukan data cleaning berupa

- Konversi menjadi huruf kecil,
- Menghapus space kosong,
- Lemmatizing setiap kata
- Penghapusan kata dengan menggunakan stopword

Hasil akhir setelah dilakukan pembersihan data adalah terdapat pengurangan jumlah data menjadi 34302

3.3.4. Sentiment Analysis Berdasarkan BERT base

Setelah features komentar bersih, dilakukan prediksi menggunakan model bert untuk mengetahui sentimen komentar pada instagram. Dengan menggunakan model BERT base diperoleh persebaran sentimen pada data post tweet sebagai berikut,

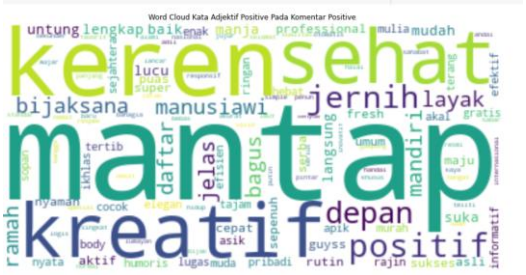


Gambar 17. *Persebaran sentimen pada Komentar Instagram*

Berdasarkan model BERT, komentar Instagram didominasi oleh sentimen netral sedangkan sentimen negatif dan positif memiliki perbedaan jumlah yang tidak terlalu jauh .

3.3.5. Analisis Kata Adjektif

Untuk mendapatkan kata-kata adjektif pada komentar instagram, dilakukan prosedur seperti *3.2.5* pada komentar instagram dan diperoleh wordcloud sebagai berikut.



Gambar 18*. Wordcloud Kata Adjektif Positif untuk Komentar Positif*

Gambar 19. *Wordcloud Kata Adjektif Negatif untuk Komentar Negatif*

Dapat dilihat *gambar 18* adalah kata-kata adjektif positif pada opini positif untuk BPJS Kesehatan seperti “keren”, “mantap”, “kreatif” sedangkan *gambar 19* adalah kata-kata adjektif negatif pada opini negatif dengan kata “sulit” dan “susah” mendominasi pada kata - kata yang negatif, dari sini dapat diduga bahwa prosedur program BPJS Kesehatan masih cukup sulit untuk dilakukan oleh masyarakat dan mungkin penyebabnya adalah kekurangan informasi masyarakat mengenai program BPJS Kesehatan. Oleh karena itu, solusi untuk BPJS kesehatan adalah melakukan penyebaran informasi sesering mungkin tentang pertanyaan yang paling sering ditanyakan oleh masyarakat Indonesia mengenai prosedur BPJS Kesehatan.

4. Kesimpulan

1. Berdasarkan hasil analisis sentiment pada data Instagram dan Twitter, hasil kinerja BPJS Kesehatan cenderung buruk dikarenakan hasil sentimen negatif lebih banyak daripada sentimen positif baik pada data Instagram dan Twitter, terutama data twitter jumlah sentimen negatifnya jauh lebih banyak dari jumlah sentimen positif.
2. Pada data instagram, kata “sulit” dan “susah” mendominasi pada komentar, dari sini diduga bahwa penyebab kinerja BPJS dinilai buruk adalah terdapat prosedur yang masih cukup sulit untuk dilakukan oleh masyarakat sedangkan pada data twitter dapat dilihat bahwa kata ‘miskin’, ‘tolol’, ‘salah’, dan ‘mahal’ banyak disebut di post tweet bertema BPJS. Sama halnya dengan data Instagram, diduga kinerja BPJS masih dinilai buruk oleh sebagian besar masyarakat.
3. Persebaran peserta BPJS terpusat di Pulau Jawa khususnya provinsi Jawa Tengah sedangkan untuk pulau lainnya peserta BPJS nya masih kurang padat.
4. Biaya tagihan BPJS tertinggi terdapat di Provinsi Papua. Jejaring adalah jenis faskes dengan biaya tagihan BPJS tertinggi. PBI APBD adalah segmen peserta dengan biaya tagihan BPJS tertinggi. TNI AU adalah kepemilikan faskes dengan biaya tagihan BPJS tertinggi.

5. Referensi

1. Devlin, J., Ming-Wei, C., Kenton, L., Kristina, T. 2018. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805v2
2. Batra, H., Narinder, S.P., Sanjay, K.S., Sonali, A. 2021. *BERT-Based Sentiment Analysis: A Software Engineering Perspective*. arXiv: 2106.02581v3
3. Hootsuite.(2022).DIGITAL 2022 INDONESIA [MOST USED SOCIAL MEDIA PLATFORM].Retrieved from https://datareportal.com/