

Analisis Sentimen Masyarakat Indonesia Terhadap Vaksin COVID-19 Secara Objektif dan Subjektif Menggunakan POS Tagging

Ryandito Diandaru¹, Muhammad Fawwaz Naabigh², Randy Zakya Suchrady³

Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung

¹13519157@std.stei.itb.ac.id

²13519206@std.stei.itb.ac.id

³13519061@std.stei.itb.ac.id

Abstrak—Pandemi COVID-19 adalah masalah yang harus dihadapi seluruh masyarakat dunia, tidak terkecuali Indonesia. Untungnya, telah ditemukan vaksin untuk mengurangi dampak dari pandemi COVID-19. Diskusi pro dan kontra dari vaksin itu sendiri pun terjadi dalam kalangan masyarakat, yang berarti ada sampel populasi yang kontra terhadap vaksinasi. Sebagai upaya membantu pengelola bidang kesehatan di Indonesia, akan dilakukan pembangunan model yang menganalisis sikap masyarakat terhadap vaksin baik secara subjektif maupun objektif dengan menggunakan POS Tagging dan eksperimen komparatif dari performa berbagai model pembelajaran mesin. Data teks dari masyarakat Indonesia mengenai vaksin yang diambil dari media sosial Twitter merupakan tweet yang diunggah pada 20 Juni 2021 - 27 Juni 2021 dan 15 Juli 2021 - 22 Juli 2021 dengan jumlah 8874 data. 6107 di antaranya dilabeli secara manual subjektivitas dan polaritasnya untuk keperluan pelatihan model. Berdasarkan skor akurasi, Multinomial Naive Bayes adalah metode yang paling baik untuk mengklasifikasi subjektivitas dengan akurasi 57%. Di sisi lain, klasifikasi polaritas untuk tweet subjektif paling baik menggunakan model Random Forest dengan akurasi 59.02%, sedangkan tweet objektif paling baik menggunakan model SVC dengan akurasi 59.56%. Hasil menunjukkan bahwa secara subjektif, masyarakat Indonesia cenderung bersentimen negatif dengan kata yang paling sering muncul di antaranya adalah 'sinovac', 'vaksin', 'enggak', 'ada', dan 'sudah', sedangkan secara objektif cenderung netral, dengan kata yang paling sering muncul di antaranya adalah 'sinovac', 'vaksin', 'pfizer', 'astrazeneca', dan 'dosis'. Dengan begitu, diharapkan dapat diambil langkah-langkah yang tepat sasaran untuk mempromosikan program vaksinasi.

Kata kunci—vaksin; POS Tagging; subjektif; objektif; sentimen;

1. Pendahuluan

1.1. Latar Belakang

Setelah berlalu lebih dari satu tahun, pandemi COVID-19 ternyata belum juga surut di tanah air. Pandemi COVID-19 yang mewabah di Indonesia sejak sekitar bulan Maret 2020 telah merenggut nyawa masyarakat Indonesia. Untungnya, vaksin sudah ditemukan untuk wabah COVID-19 ini. Vaksin sudah diproduksi secara massal oleh sejumlah perusahaan serta secara umum masyarakat di seluruh dunia bisa mendapatkannya secara gratis dari pemerintahannya.

Sosial media Twitter dapat digunakan sebagai sumber pengambilan opini masyarakat Indonesia terhadap vaksin COVID-19. Untuk itu, pemetaan sentimen masyarakat terhadap vaksin yang beredar bisa dilakukan, dengan harapan dapat memberikan wawasan lebih untuk badan yang bertanggungjawab dalam penanganan pandemi COVID-19 agar dapat dirumuskan strategi sosialisasi yang efektif kepada masyarakat.

1.2. Tujuan dan Manfaat

1.2.1. Tujuan

Penelitian ini dilakukan untuk membangun model dengan menggunakan *machine learning* dari data Twitter tentang vaksin untuk mengklasifikasi subjektivitas sebuah *tweet*, yakni objektif atau subjektif. Model juga dapat mengklasifikasi sentimen *tweet*, yakni positif, negatif, atau netral.

1.2.2. Manfaat

Klasifikasi subjektivitas diperlukan agar diketahui apakah sebuah *tweet* bersifat subjektif/objektif. Sementara itu, analisis sentimen bermanfaat agar diketahui sebuah *tweet* bernada pro atau kontra terhadap vaksin. Dengan demikian, para pemangku kepentingan diharap dapat mengetahui bagaimana pandangan masyarakat mengenai vaksin dan kejadian apa saja yang berhubungan dengan vaksin yang ditemui masyarakat, sehingga dapat diambil langkah-langkah strategis demi meningkatkan minat masyarakat untuk divaksin.

1.3. Batasan yang Digunakan

- Model tidak memperhitungkan data hasil *scraping* yang menggunakan bahasa selain Bahasa Indonesia.
- Perwakilan masyarakat Indonesia yang aktif di media sosial Twitter diasumsikan sudah mewakili pendapat seluruh populasi.
- Pelabelan data secara manual dilakukan secara mandiri.
- Pustaka Twitter API yang digunakan tidak berbayar.

1.4. Kajian Serupa

Beberapa penelitian terdahulu telah membahas topik serupa, di antaranya adalah:

- *An analysis of COVID-19 vaccine sentiments and opinions on Twitter* oleh Yousefinaghani, dkk (2021). Penelitian tersebut meneliti tentang kemauan masyarakat dunia untuk divaksin, dengan menganalisis 4,5 juta *tweets* yang diambil dari Twitter dengan menggunakan Valence Aware Dictionary and Sentiment Reasoner (VADER) untuk menentukan polaritas dari *tweet*. Hasil menunjukkan bahwa polaritas positif lebih dominan.
- *Analyzing the attitude of Indian citizens towards COVID-19 vaccine – A text analytics study* oleh Praveen, dkk (2021). Penelitian tersebut menganalisis sentimen masyarakat India terhadap diadakannya program vaksinasi massal oleh pemerintah India melalui postingan yang diekstrak dari internet. Karena analisis sentimen yang dilakukan dilakukan pada level dokumen, kalimat, dan frasa kurang bisa menunjukkan alasan mengapa terdapat masyarakat India yang masih enggan menerima dosis vaksin, penelitian dilanjutkan dengan analisis lebih lanjut untuk mengetahui alasan sebagian sampel bersentimen negatif. Hasil menunjukkan bahwa sebanyak 47% masyarakat India netral terhadap vaksin, hampir 17% negatif dengan alasan yang dominan yaitu kekhawatiran akan reaksi alergi dan kesehatan.

1.5. Celah Penelitian

Berdasarkan kajian penelitian yang telah dilakukan, ditemukan beberapa kesempatan pengolahan data yang belum dilakukan terutama pada pembedaan *tweet* yang bersifat objektif dan subjektif. Dengan tidak dilakukan klasifikasi antara *tweet* bernada subjektif dengan objektif, bias berpotensi ditimbulkan karena klaim kurang bersifat atas dasar ilmiah. Meskipun *tweet* objektif kurang menggambarkan pendapat masyarakat mengenai vaksin COVID-19, *tweet-tweet* tersebut lebih menggambarkan argumen-argumen yang didasari riset oleh masyarakat yang kompeten untuk membahas vaksin COVID-19. Penelitian ini

didasari oleh niat untuk memperluas wawasan yang didapatkan dengan meneliti lebih dalam akan signifikansi dari objektivitas suatu teks mengenai vaksin oleh masyarakat.

Selain itu, temuan lain dalam kajian penelitian sebelumnya belum menekankan analisis objektif pada jenis vaksin yang telah beredar, melainkan hanya vaksin secara umum. Dengan demikian, diinginkan tambahan wawasan lebih lanjut jika terdapat sampel populasi bersentimen negatif yang berkemungkinan hanya disebabkan oleh ketidakcocokan subjektif dengan jenis vaksin.

1.6. Usulan Solusi

Untuk menambah solusi dari penelitian-penelitian yang sudah pernah dilakukan, ditawarkan sebuah solusi berupa model yang juga mengklasifikasi antara *tweet* subjektif dan objektif untuk data teks tentang vaksin yang berasal dari Twitter.

2. Studi Literatur

2.1. Penambahan Opini dan Analisis Sentimen

Dalam bidang *Natural Language Processing*, *Sentiment Analysis* (SA) merupakan metode yang digunakan untuk menambang informasi atau opini pada topik tertentu. Salah satu konsep dasar pada SA adalah mengidentifikasi kalimat yang dianalisis berdasarkan subjektivitas dan objektivitas (Saberi dan Saad, 2017). Pengidentifikasian ini penting untuk dilakukan, sebab kalimat yang dianalisis bisa saja berisi fakta-fakta objektif yang bertentangan dengan sifat opini itu sendiri, yakni subjektif. Teknik yang digunakan dalam pengidentifikasian ini dikenal dengan istilah *opinion mining*, yaitu teknik yang digunakan untuk mendeteksi dan mengekstrak informasi yang bersifat subjektif dari sebuah kalimat (Padmaja dan Fatima, 2013). Setelah subjektivitas kalimat teridentifikasi, langkah selanjutnya adalah mengklasifikasikan apakah kalimat tersebut bersentimen positif, negatif, atau netral. Sentimen netral dilabelkan pada kalimat yang bukan termasuk pada sentimen positif atau negatif.

2.2. POS Tagging

Part-of-Speech (POS) *Tagging* merupakan salah satu pemrosesan data teks pada sektor leksikal dengan cara memberikan label pada kata yang terdapat di dalam teks (Purwarianti, 2014). Terdapat dua jenis/kelas kata yang terklasifikasi di dalam POS *tagging*: *Open class words* dan *Closed class words*. *Open class words* merupakan kelas kata yang melingkupi kata benda, kata kerja, kata sifat, dan jenis kata yang dapat terus bertambah anggotanya. Sedangkan, *Closed class words* merupakan kelas kata yang berisikan kata-kata yang relatif telah ditetapkan dan cenderung jumlahnya terbatas. *Closed class words* berisikan kata-kata seperti kata hubung, kata ganti, kata keterangan posisi, dan kata lainnya.

Dalam melakukan POS Tagging, terdapat tiga metode, yakni *Rule-based Tagger*, *Stochastic Tagger*, dan *Transformation-based Tagger* (Saefulloh, 2017). *Rule-based Tagger* merupakan metode pendekatan *top-down* dengan mendefinisikan aturan baku tertentu. Pada *Stochastic Tagger*, digunakan data *training* untuk menentukan probabilitas *tag* sebuah kata. Untuk metode *Transformation-based Tagger*, akan dipadukan kedua cara sebelumnya dengan menjadikan *knowledge* dari korpus data *training* sebagai aturan atau *rule*. Terdapat beberapa algoritma yang secara umum dipakai untuk melakukan POS Tagging, yakni Hidden Markov Model (HMM) yang menggunakan probabilitas sebuah kata dapat dilabeli dengan label jenis kata tertentu, Decision Tree atau pohon keputusan, dan Neural Network yang menghasilkan fungsi pembeda dari pelatihan modelnya.

3. Metode

3.1. Proses Pengumpulan Data

3.1.1. *Scraping Data Tweet*

Dokumen-dokumen teks merupakan hasil ekstraksi dari Twitter melalui API Twitter dengan menggunakan pustaka Tweepy versi 3.10.0 (dokumentasi dapat diakses di <https://docs.tweepy.org/en/v3.10.0/>) untuk *tweet* pada 20 Juni 2021 - 27 Juni 2021 dan 15 Juli 2021 - 22 Juli 2021. Filter yang digunakan untuk ekstraksi *tweet* adalah sebagai berikut.

1. Vaksin -filter:retweets
2. astrazeneca OR sinovac OR sinopharm OR coronavac -filter:retweets

Proses *scraping* menghasilkan 8874 entri *tweet*, 6107 di antaranya dilabeli secara manual untuk keperluan *training* model. Namun, dari 6107 data tersebut, sebanyak 1924 data ditemukan ditulis dalam bahasa daerah, sehingga tidak dilabeli. Profil data yang telah dilabeli secara manual adalah sebagai berikut.

Tabel 3.1 Profil data yang telah dilabeli

	Positif	Negatif	Netral	Subjektif	Objektif
Jumlah	1145	898	2141	2327	1857

Adapun atribut dataset mentah adalah sebagai berikut.

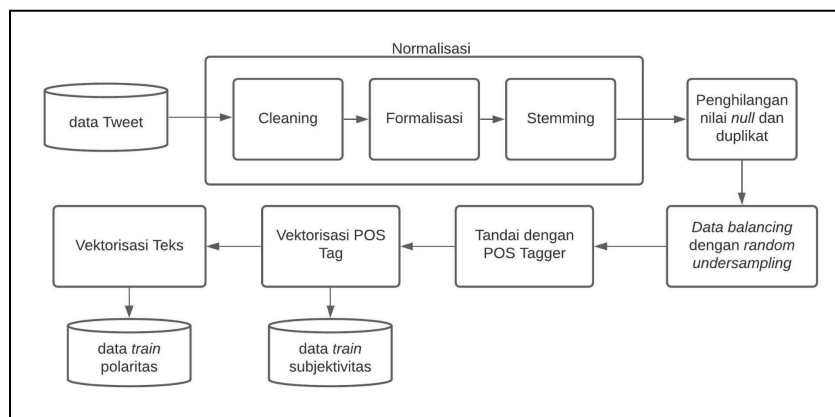
1. id - *Index* : identitas sebuah baris data
2. text (full_text) - *String* : isi dari *tweet* terkait

3.1.2. *Data Training Model POS Tagger*

Training model POS Tagger akan menggunakan dataset kalimat-kalimat berbahasa Indonesia yang sudah di-tag secara manual oleh Manurung, dkk (2015) yang dapat diakses pada github.com/famrashel

3.2. Pemrosesan Data Teks

Pra-proses digambarkan pada diagram di bawah ini.



Gambar 3.1 Diagram pra-pemrosesan data

Pada tahap normalisasi, *cleaning* digunakan untuk merubah seluruh data menjadi huruf kecil, menghilangkan tanda baca, dan karakter spesial. Formalisasi dilakukan untuk merubah kata yang dianggap sebagai singkatan atau *slang* menjadi kata baku, proses ini dilakukan merujuk pada dataset “Kamus Alay - Colloquial Indonesian Lexicon” oleh Salsabila, dkk (2018). *Stemming* dilakukan untuk merubah kata yang telah baku menjadi kata dasar. Setelah proses normalisasi, didapat jumlah kata atau token sebanyak 7735 kata yang awalnya berjumlah 17540 kata.

Adapun fitur data latih setelah dilakukan normalisasi adalah sebagai berikut.

1. id - *Index* : identitas sebuah baris data
2. text (full_text) - *String* : isi dari *tweet* terkait
3. norm - *String* : isi dari *tweet* terkait setelah dilakukan normalisasi
4. formal - *String* : isi dari *tweet* terkait yang telah dinormalisasi dan dilakukan perubahan kata yang tidak formal, singkatan dan slang, menjadi kata yang formal
5. subjektif - *Categoric* : subjektivitas dari *tweet* terkait
6. posneg - *Categoric* : sentimen dari *tweet* terkait (positif, negatif, atau netral)

Sebelum dilakukan vektorisasi, setiap kata akan diubah menjadi tag POS yang mewakili kata tersebut. Fitur data latih sebelum vektorisasi adalah sebagai berikut.

1. tagonly - *String* : berisi POS tag setiap kata dari kalimat secara berurutan
2. subjektif - *Categoric* : subjektivitas dari *tweet* terkait

Sementara itu, fitur data latih pengklasifikasian sentimen sebelum vektorisasi adalah sebagai berikut.

1. formal - *String* : isi dari *tweet* terkait yang telah dinormalisasi dan dilakukan perubahan kata yang tidak formal, singkatan dan slang, menjadi kata yang formal
2. posneg - *Categoric* : sentimen dari *tweet* terkait (positif, negatif, atau netral)

3.3. Desain Eksperimen

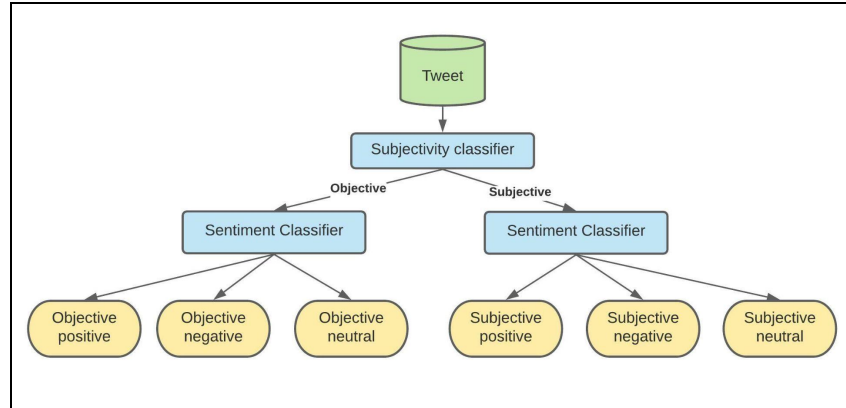
Model akan mengklasifikasi setiap *tweet* menjadi enam kelas tergantung subjektivitas dan polaritasnya.

		Tone	
		Subjective	Objective
Polarity	Negative	1	2
	Positive	3	4
	Neutral	5	6

Gambar 3.2 Matriks pengklasifikasian data, modifikasi dari Schumaker, dkk (2016)

3.3.1. Desain Arsitektur Model

Model akan mengklasifikasikan subjektivitas terlebih dahulu. Teknik yang digunakan adalah teknik *rule based* dan pembuatan model *classifier*, yakni dengan Multinomial Naive Bayes, SVC, Random Forest, Decision Tree, dan Logistic Regression. Skor yang diambil dari tiap teknik/model adalah akurasi. Kemudian, data akan diklasifikasikan polaritasnya. Eksperimen pengklasifikasian polaritas *tweet* dilakukan dengan pembuatan model *classifier* yang sama dengan pengklasifikasian subjektivitas.



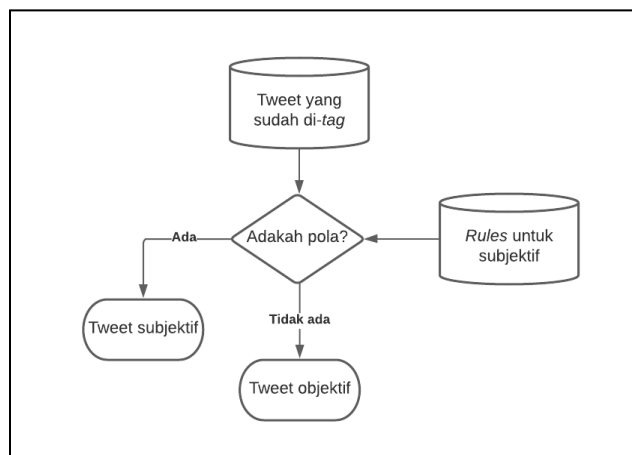
Gambar 3.3 Diagram alur keseluruhan sistem

Adapun rule yang digunakan untuk menentukan subjektivitas adalah sebagai berikut. Rules yang digunakan merupakan penyesuaian dari rule yang digunakan oleh Rozi, dkk. (2012, hal. 5).

Tabel 3.2 Rule pengklasifikasian kalimat objektif, Rozi, dkk. (2012, hal. 5)

Rule	Contoh	Rule	Contoh
RB JJ	Sungguh bagus, terlampau bagus	JJ RB	Indah sekali, bagus sekali
RB VB	Semoga membawa hikmah	VB JJ	Membuat bingung
NN JJ	Bukunya bagus, pakaiannya rapi	NEG JJ	Tidak seindah, tidak semudah
NN VB	Pelajarannya membosankan	NEG VB	Tidak mengerti, tidak memahami
JJ VB	Mudah difahami, gampang dimaafkan	PRP VB	Saya menyukai, kita suka
VB VB	Membuat merinding, membuat pusing	MD VB	Perlu mengambil referensi, Perlu dikembangkan

Skema pengklasifikasian subjektivitas dengan *rule* adalah sebagai berikut.

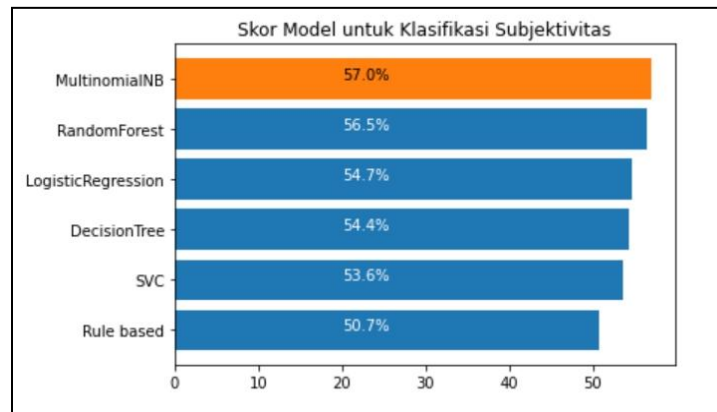


Gambar 3.4 Alur pengklasifikasian tweet subjektif/objektif dengan rules

4. Hasil dan Analisis Pengujian

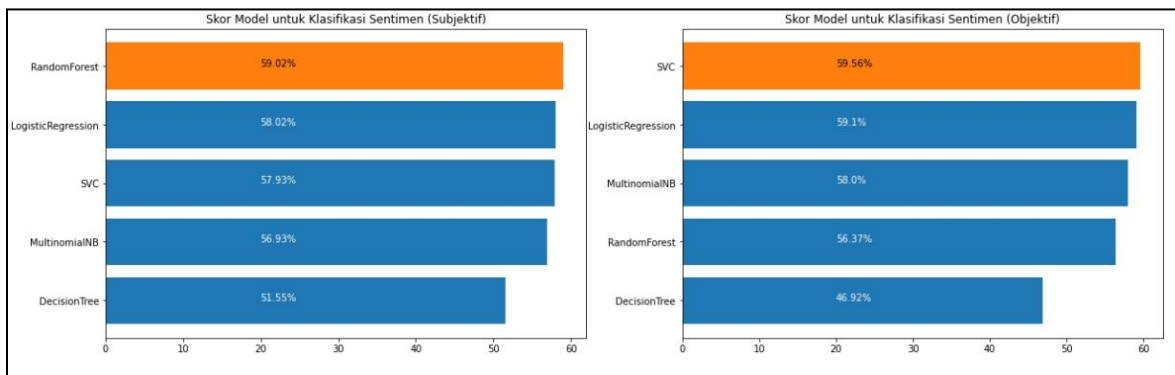
4.1. Hasil

Evaluasi dilakukan menggunakan metode Cross Validation dengan 5 *fold* dan diambil rata-rata dari seluruh skor individual tiap *fold* sebagai akurasi akhir. Eksperimen klasifikasi subjektivitas *rule based* adalah pengecualian karena metode klasifikasi ini tidak memerlukan model. Gambar di bawah ini menunjukkan bahwa berdasarkan skor akurasi, metode yang paling baik adalah dengan model Multinomial Naive Bayes.



Gambar 4.1 Skor akurasi pengklasifikasian subjektivitas dengan POS Tagging

Evaluasi klasifikasi polaritas dilakukan dengan ukuran skor yang sama, yakni akurasi, tetapi Cross Validation menggunakan 10 *fold*. Gambar di bawah menunjukkan berdasarkan skor akurasi, metode terbaik untuk subjektif adalah Random Forest dan Support Vector Classifier untuk objektif.



Gambar 4.2 Skor akurasi pengklasifikasian sentimen untuk tweet subjektif dan objektif

4.2. Analisis

Berikut adalah sampel hasil klasifikasi dua tahap yang telah dilakukan. O menandakan objektif dan S menandakan subjektif, kemudian N menandakan sentimen negatif, T menandakan netral, dan P menandakan positif. Sampel nomor satu menunjukkan hasil klasifikasi yang benar dalam subjektivitas dan polaritas, nomor dua menunjukkan hasil klasifikasi yang benar dalam subjektivitas tetapi salah dalam polaritas, nomor tiga menunjukkan hasil klasifikasi yang salah dalam subjektivitas tetapi benar dalam polaritas, dan nomor empat menunjukkan hasil klasifikasi yang salah dalam subjektivitas dan polaritas.

Tabel 4.1 Contoh kesalahan dalam pengklasifikasian

No	text	Subjektif	Posneg
1	Update 12 jam after vaksin sinovac dosis ke 2 ini, badan gue lemes dan pegel banget huff.	O	N
2	@RestyCayaaah AKU GK PERCAYA VAKSIN AKU GK PERCAYA COVID. PERSETAN DENGAN JOKOWI 😡	S	T
3	@_epiphiany Wei kalau pfizer/sinovac mmg ngantuk gilerrrr	S	N
4	@AndiDwan_ @unmagnetism Normal kok, ane dapet vaksinasi sinovac jg gitu, pegelnya gak lama & gak lebih ngilu gitu. Ini yg ane alami jg kok 👍	S	N

Kasus mengenai sebuah tweet terklasifikasikan dengan benar dapat menjadi akibat dari padanan kata pada proses training yang cocok dengan kalimat *tweet* yang akan diklasifikasi. Sementara itu, kesalahan pada pengklasifikasian dapat terjadi akibat dari beberapa hal, yaitu

1. *Tweet* ditulis dalam Bahasa Indonesia, tetapi mengandung istilah asing atau istilah daerah.
2. Terdapat istilah slang yang tidak terdaftar pada dataset yang dirujuk untuk formalisasi.
3. *Tweet* bersifat sarkasme, sehingga memiliki makna sebaliknya.
4. Terdapat kesalahan tulisan (*typo*) pada *tweet*.

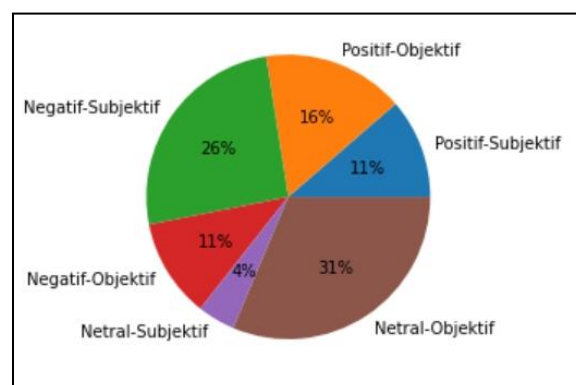
Tabel 4.2 Contoh kasus kesalahan dalam normalisasi

Sebelum normalisasi	Setelah normalisasi
@Miangkey bkn x vaksin, x dapat2 vaksin nya	bukan kali vaksin kali vaksin nya

Terlihat bahwa karakter “x” dinormalisasi menjadi “kali”, padahal “x” pada data teks tersebut bermakna “tidak”. Sementara itu, hasil POS Tagging kalimat tersebut adalah sebagai berikut

Bukan : NEG, kali : NND, vaksin : NN, kali : NN, kali : NN, nya : PRP

Hasil normalisasi yang kurang optimal juga akan menghasilkan hasil POS Tagging yang juga kurang optimal, beberapa kata terdeteksi sebagai NN (kata benda) padahal berarti lain.



Gambar 4.3 Grafik persentase dari tiap kelas

Berdasarkan grafik di atas, dapat dilihat bahwa *tweet* subjektif adalah sekitar sebesar 41% dari seluruh data tes, yang berarti *tweet* objektif adalah sekitar sebesar 59%. Besar persentase masing-masing kombinasi kelas antara positif-negatif-netral dan subjektif-objektif beserta frekuensi lima kata yang paling sering muncul adalah sebagai berikut.

Tabel 4.3 Persentase kombinasi kelas dan lima kata terbanyak yang paling sering muncul beserta jumlah kemunculannya

Kelas	Persentase	Frekuensi kata terbanyak
Positif-Objektif	16%	sinovac: 409, vaksin: 375, aku: 161, enggak: 99, sudah: 90
Negatif-Objektif	11%	sinovac: 267, vaksin: 181, enggak: 91, gue: 72, aku: 56
Netral-Objektif	31%	sinovac: 663, vaksin: 629, pfizer : 193, astrazeneca : 184, dosis: 112
Positif-Subjektif	11%	vaksin: 347, sinovac: 255, sudah: 73, enggak: 73, ada: 57
Negatif-Subjektif	26%	sinovac: 669, vaksin: 337, enggak: 154, ada: 135, sudah: 107
Netral-Subjektif	4%	vaksin: 118, sinovac: 112, saya: 28, saja: 27, enggak: 27

Hasil kuantitatif menunjukkan bahwa secara subjektif, masyarakat Indonesia lebih banyak bersentimen negatif terhadap vaksin COVID-19 daripada bersentimen positif maupun netral. Dapat dilihat bahwa *tweet* bernada positif, negatif, maupun netral banyak memiliki kata kunci ‘vaksin’ dan ‘sinovac’. Kata-kata yang berfrekuensi tinggi dianggap sebagai akibat dari jenis vaksin yang memang sedang beredar di Indonesia. Dari hasil eksperimen dapat ditarik sebuah hipotesis bahwa sentimen negatif yang dominan oleh masyarakat Indonesia berkaitan dengan kata-kata yang banyak disebutkan pada *tweet-tweet* sentimen itu sendiri.

Di sisi lain, secara objektif, masyarakat Indonesia lebih dominan bersentimen netral daripada positif maupun negatif. Namun, tidak menutup kemungkinan bahwa hasil yang didapat dipengaruhi oleh kalimat objektif yang bersifat menyatakan fakta dan cenderung netral. Berbeda dengan *tweet* yang diklasifikasikan sebagai subjektif, mayoritas dari kelas objektif, yakni objektif-netral banyak menyebutkan kata kunci yang merupakan jenis-jenis vaksin COVID-19.

5. Simpulan

Hasil eksperimen menunjukkan bahwa *classifier* yang paling baik untuk persoalan subjektivitas adalah model Multinomial Naive Bayes. Di sisi lain, klasifikasi sentimen paling baik menggunakan model Random Forest untuk *tweet* yang sudah diklasifikasikan sebagai *tweet* subjektif dan model SVC untuk *tweet* yang sudah diklasifikasikan sebagai *tweet* objektif. *Tweet* yang terklasifikasi sebagai subjektif menunjukkan bahwa masyarakat Indonesia dominan bersentimen negatif terhadap vaksin, dengan lima kata yang paling banyak disebutkan untuk kelas negatif-subjektif adalah ‘sinovac’, ‘vaksin’, ‘enggak’, ‘ada’, dan ‘sudah’. Sedangkan untuk *tweet* objektif, masyarakat Indonesia lebih dominan bersentimen netral dengan lima kata yang paling banyak disebutkan untuk kelas netral-objektif adalah ‘sinovac’, ‘vaksin’, ‘pfizer’, ‘astrazeneca’, dan ‘dosis’. Dengan informasi ini, diharapkan badan yang bertanggung jawab terhadap penanganan pandemi COVID-19 di Indonesia dapat mendapatkan wawasan yang lebih dan dapat merumuskan strategi vaksinasi untuk masyarakat Indonesia.

DAFTAR PUSTAKA

- Dinakaramani, Rashel, dkk. 2014. “*Designing an Indonesian Part of Speech Tagset and Manually Tagged Indonesian Corpus*”. cs.ui.ac.id. Diakses pada 28 Juni 2021, dari <https://bahasa.cs.ui.ac.id/postag/downloads/Designing%20an%20Indonesian%20Part%20of%20speech%20Tagset.pdf>
- Padmaja, Fatima. 2013. “*Opinion Mining and Sentiment Analysis - An Assessment of Peoples' Belief: A Survey*”. researchgate.net. Diakses pada 28 Juni 2021, dari https://www.researchgate.net/publication/276196657_Opinion_Mining_and_Sentiment_Analysis_-_An_Assessment_of_Peoples'_Belief_A_Survey
- Praveen, dkk. 2021. “Analyzing the attitude of Indian citizens towards COVID-19 vaccine - A text analytics study” Diakses pada 28 Juni 2021, dari <https://pubmed.ncbi.nlm.nih.gov/33714134/>
- Purwarianti. 2014. “*POS Tagger Ayu Purwarianti*”. slideshare.net. Diakses pada 28 Juni 2021, dari <https://www.slideshare.net/ayupurwarianti/pos-tagger-37881733>
- Rozi, Pramono, dkk. 2013. “Implementasi *Opinion Mining* (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi”. neliti.com. Diakses pada 28 Juni 2021, dari <https://media.neliti.com/media/publications/69248-ID-implementasi-opinion-mining-analisis-sen.pdf>
- Saberi, Saad. 2017. “*Sentiment Analysis or Opinion Mining: A Review*”. researchgate.net. Diakses pada 28 Juni 2021, dari https://www.researchgate.net/publication/320762707_Sentiment_Analysis_or_Opinion_Mining_A_Review
- Saeffulloh. 2017. “*Part of Speech Tagger untuk Bahasa Indonesia untuk Conditional Random Field*”. coggle.it. Diakses pada 29 Juni 2021, dari <https://coggle.it/diagram/XJXasQTz8WldUTgI/t/part-of-speech-tagger-untuk-conditional-random-field-crf>
- Salsabila, Winatmoko, dkk. 2018. “Colloquial Indonesian Lexicon”. ieee.org. Diakses pada 24 Juli 2021, dari <https://ieeexplore.ieee.org/abstract/document/8629151>.
- Schumaker, dkk. 2016. “Predicting wins and spread in the Premier League using a sentiment analysis of twitter” Diakses pada 28 Juni 2021, dari https://www.researchgate.net/publication/303892471_Predicting_wins_and_spread_in_the_Premier_League_using_a_sentiment_analysis_of_twitter
- Yousefinaghani, Dara, dkk. 2021. “An analysis of COVID-19 vaccine sentiments and opinions on Twitter”. Diakses pada 28 Juni 2021, dari [https://www.ijidonline.com/article/S1201-9712\(21\)00462-8/fulltext](https://www.ijidonline.com/article/S1201-9712(21)00462-8/fulltext)