

**IMPLEMENTING NATURAL LANGUAGE PROCESSING FOR SENTIMENT  
ANALYSIS OF PUBLIC SERVICE DISCOURSE ON TWITTER**

**PENERAPAN NATURAL LANGUAGE PROCESSING UNTUK ANALISIS  
SENTIMEN TERHADAP LAYANAN PUBLIK DI MEDIA SOSIAL TWITTER**

**Dhimas Tribuana<sup>1\*</sup>, Usman<sup>2</sup>, Dayanti<sup>3</sup>**

Pascasarjana, Universitas Komputer Indonesia, Bandung<sup>1</sup>

Bisnis Digital, Politeknik Negeri Bombana, Bombana<sup>2</sup>

Fakultas Teknik, Universitas Patria Artha, Makassar<sup>3</sup>

[dhimas.75423008@mahasiswa.unikom.ac.id](mailto:dhimas.75423008@mahasiswa.unikom.ac.id)<sup>1</sup>, [usmanfkumi@gmail.com](mailto:usmanfkumi@gmail.com)<sup>2</sup>,

[dayanti.fattah@gmail.com](mailto:dayanti.fattah@gmail.com)<sup>3</sup>

---

**ABSTRACT**

*This study aims to develop and evaluate a sentiment classification model of public opinion on BPJS Ketenagakerjaan services using Twitter data. An experimental quantitative approach was applied, consisting of five stages: data collection, text preprocessing, data labeling, feature extraction and model training, and performance evaluation. Data were collected via Twitter API using keywords related to BPJS Ketenagakerjaan services, yielding 4,357 tweets. After preprocessing and labeling, a clean dataset of 4,102 tweets was categorized into three sentiment classes: positive, negative, and neutral. Two text representation methods were employed, namely TF-IDF and IndoBERTweet embeddings, and tested with machine learning algorithms such as Naive Bayes, SVM, Random Forest, and Logistic Regression. The results show that the IndoBERTweet + Logistic Regression model achieved the best performance with an accuracy of 86.5% and an F1-score of 85.6%. These findings highlight the effectiveness of pretrained language models in understanding informal Indonesian language on social media. The developed model has potential as a real-time public opinion monitoring system to improve BPJS Ketenagakerjaan service quality. Future research is recommended to explore sarcasm handling and multi-platform datasets to enhance model generalization.*

**Keywords:** Public sentiment, BPJS Ketenagakerjaan, IndoBERTweet, text classification, social media

**ABSTRAK**

Penelitian ini bertujuan untuk mengembangkan dan menguji model klasifikasi sentimen publik terhadap layanan BPJS Ketenagakerjaan berdasarkan data media sosial Twitter. Metode yang digunakan adalah pendekatan kuantitatif eksperimental yang terdiri dari lima tahap, yaitu pengumpulan data, pra-pemrosesan teks, pelabelan data, ekstraksi fitur dan pelatihan model, serta evaluasi performa. Data dikumpulkan melalui API Twitter menggunakan kata kunci terkait layanan BPJS Ketenagakerjaan dengan total 4.357 tweet. Setelah pra-pemrosesan dan pelabelan, data bersih sebanyak 4.102 tweet diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Dua pendekatan representasi teks digunakan, yaitu TF-IDF dan IndoBERTweet embedding, yang kemudian diuji dengan algoritma pembelajaran mesin seperti *Naive Bayes*, *SVM*, *Random Forest*, dan *Logistic Regression*. Hasil penelitian menunjukkan bahwa model IndoBERTweet + Logistic Regression memberikan performa terbaik dengan akurasi 86,5% dan F1-score 85,6%. Temuan ini menegaskan efektivitas *pretrained language models* dalam memahami bahasa Indonesia informal di media sosial. Model yang dikembangkan memiliki potensi untuk digunakan sebagai sistem pemantauan opini publik secara *real-time* guna meningkatkan kualitas layanan BPJS Ketenagakerjaan. Penelitian selanjutnya disarankan mengeksplorasi penanganan sarkasme dan dataset multi-platform untuk meningkatkan generalisasi model.

**Kata Kunci:** Sentimen publik, BPJS Ketenagakerjaan, IndoBERTweet, klasifikasi teks, media sosial.

*This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).*

Artikel ini adalah artikel akses terbuka yang didistribusikan di bawah ketentuan Lisensi Creative Commons Attribution 4.0 International (CC BY 4.0).



## PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan besar dalam berbagai sektor, termasuk sektor pelayanan publik. Pemerintah Indonesia secara aktif mendorong pemanfaatan teknologi digital untuk meningkatkan kualitas layanan kepada masyarakat, termasuk melalui kehadiran Badan Penyelenggara Jaminan Sosial (BPJS) Ketenagakerjaan yang bertugas memberikan perlindungan sosial ekonomi kepada tenaga kerja (Kemenko PMK, 2022). Seiring dengan meningkatnya kesadaran masyarakat terhadap hak-hak ketenagakerjaan, ekspektasi terhadap pelayanan yang cepat, responsif, dan transparan juga semakin tinggi. Namun demikian, kualitas layanan publik sering kali masih menjadi keluhan utama masyarakat, khususnya di media sosial, yang kini menjadi kanal ekspresi opini publik secara langsung dan terbuka (Hidayat & Wenggi, 2022).

Media sosial, terutama Twitter, telah menjadi ruang diskusi dinamis yang mencerminkan aspirasi, kritik, hingga pujian terhadap instansi pemerintah. Penelitian sebelumnya menyebutkan bahwa Twitter dapat digunakan sebagai sumber informasi yang representatif untuk memantau persepsi publik secara *real-time* (Naraswati et al., 2021; Pak & Paroubek, 2010). Dalam konteks BPJS Ketenagakerjaan, isu seperti keterlambatan pencairan dana JHT, kendala dalam proses klaim, atau kurangnya respons petugas layanan sering kali muncul dalam bentuk unggahan atau keluhan netizen (Maharani & Mandira, 2022). Namun, data-data ini belum banyak dimanfaatkan secara sistematis sebagai alat evaluasi kinerja layanan. Oleh karena itu, dibutuhkan pendekatan berbasis data yang dapat mengolah informasi dari media sosial secara masif dan otomatis, salah satunya dengan pendekatan *Natural Language Processing* (NLP).

NLP merupakan cabang dari kecerdasan buatan (AI) yang memungkinkan komputer untuk memahami, menginterpretasikan, dan menghasilkan bahasa manusia secara efektif (Jurafsky & Martin, 2023). Dalam konteks analisis sentimen, NLP dapat digunakan untuk mengklasifikasikan opini publik menjadi kategori positif, negatif, atau netral. Teknik ini telah digunakan secara luas dalam berbagai studi seperti analisis kepuasan pelanggan (Rani & Kumar, 2017), pemantauan opini politik (Mukti et al., 2023), dan penilaian reputasi merek (Liu, 2012). Dalam Bahasa Indonesia, NLP menghadapi tantangan tersendiri, seperti kompleksitas struktur kalimat dan minimnya korpus beranotasi, namun berbagai inisiatif seperti penggunaan pustaka Sastrawi dan IndoBERT telah membantu meningkatkan akurasi pemrosesan teks lokal (Adhiatma & Qoiriah, 2022; Wilie et al., 2020).

Berbagai studi juga menunjukkan bahwa kombinasi NLP dan algoritma machine learning dapat menghasilkan klasifikasi sentimen yang akurat pada data Twitter. Model seperti *Naive Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* terbukti mampu mengenali pola-pola linguistik yang mencerminkan emosi pengguna (Aji et al., 2022; Madhoushi et al., 2015). Penelitian oleh Kurniawan et al. (2023) menemukan bahwa penggunaan SVM pada data Bahasa Indonesia dari Twitter menghasilkan akurasi di atas 80% untuk klasifikasi sentimen (Kurniawan et al., 2023). Hal ini menunjukkan potensi besar analisis sentimen sebagai alat bantu pengambilan keputusan dalam kebijakan publik, termasuk dalam mengukur efektivitas dan kualitas layanan BPJS Ketenagakerjaan berdasarkan persepsi masyarakat.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan metode NLP dalam menganalisis sentimen masyarakat terhadap layanan BPJS Ketenagakerjaan yang diekspresikan melalui Twitter. Data yang digunakan merupakan data real hasil crawling menggunakan API Twitter dengan kata kunci tertentu terkait BPJS Ketenagakerjaan. Penelitian ini diharapkan dapat memberikan kontribusi terhadap upaya peningkatan kualitas pelayanan publik melalui pendekatan berbasis data, sekaligus memberikan bukti bahwa media sosial dapat menjadi instrumen monitoring yang kredibel dan real-time dalam ekosistem pelayanan pemerintah di era digital.

## TINJAUAN PUSTAKA

### Analisis Sentimen pada Layanan BPJS

Beberapa penelitian telah menganalisis sentimen publik terkait layanan BPJS melalui Twitter. Akbar & Habibi (2023) mengembangkan metode hibrida (*lexicon-based + Naive Bayes*) untuk menganalisis 4.154 tweet terkait JHT (Jaminan Hari Tua) dan BPJS Ketenagakerjaan, mencapai akurasi hingga 95 % (Rizky Fauzi Akbar & Habibi, 2023). Nada et al. (2023) membandingkan *Naive Bayes* dan SVM menggunakan data twitter sejak 2019–2021, dan menemukan SVM Kernel RBF unggul dengan akurasi 97,1 % (Nada et al., 2023). Studi-studi ini menunjukkan efektivitas NLP dalam menangkap sentimen spesifik terhadap program ketenagakerjaan BPJS.

### Metode Pra-pemrosesan dan Klasifikasi

Pra-pemrosesan menjadi elemen penting dalam analisis sentimen Indonesia. Kardian & Gustiana (2021) menerapkan *case folding, filtering, negation handling, stopword removal*, dan *stemming* sebagai tahapan awal, menghasilkan akurasi ~71 % (Kardian & Gustiana, 2021). Sementara itu, penelitian BPJS Kesehatan oleh Andrianti (2018) menggunakan *Naive Bayes, emoticon handling*, dan tokenisasi untuk mencapai akurasi 88 % (Andrianti, 2018). Model seperti ini dapat diadaptasi untuk BPJS Ketenagakerjaan.

### Teknik *Ensembl* dan Hybrid

Anam et al. (2022) mengombinasikan SVM, Adaboost, PSO, dan *Genetic Algorithm* untuk analisis sentimen BPJS, mencapai akurasi hingga 92 % (Anam et al., 2022). Ini memperlihatkan bahwa teknik gabungan dapat meningkatkan performa, terutama saat data sosial media beragam dan banyak *noise*.

### Model *Deep Learning* dan *Transfer Learning* untuk Twitter Bahasa Indonesia

Model deep learning mulai diterapkan dalam NLP Bahasa Indonesia. IndoBERTweet (Koto et al., 2021) adalah model besar yang dilatih khusus pada korpus Twitter Indonesia, dengan hasil signifikan pada berbagai dataset Tweet (Koto et al., 2021). Sedangkan Azhar & Khodra (2021) melakukan *fine-tuning* IndoBERT untuk domain keuangan, dan menemukan peningkatan signifikan pada tugas *sentiment analysis domain*-spesifik (Azhar & Khodra, 2021). Pendekatan ini menjanjikan untuk studi tentang BPJS, mengingat konteks spesifik layanan publik.

### Perspektif Evaluasi Layanan Publik melalui Layanan Sosial Media

Penggunaan analisis sentimen untuk evaluasi layanan publik telah menjadi tren global. Pak & Paroubek (2010) mendemonstrasikan bahwa media sosial, khususnya Twitter, efektif untuk *opinion mining* dan *tracking* opini (Pak & Paroubek, 2010). Naraswati et al. (2020) juga menyoroti data Twitter sebagai alat pemantauan sosial publik secara *real-time* (Naraswati et al., 2021). Di konteks layanan kesehatan, Girsang et al. (2022) menemukan sekitar 60 % opini negatif untuk BPJS Kesehatan dengan model LDA + NLP (Girsang et al., 2023). Diambil bersama, penelitian-penelitian ini menegaskan potensi implementasi analisis sentimen terhadap BPJS Ketenagakerjaan.

### Tantangan dalam Bahasa Indonesia

Pengolahan teks Bahasa Indonesia menghadapi tantangan seperti struktur sintaksis, variasi dialek, slang, dan keterbatasan korpus anotasi (Adhiatma & Qoiriah, 2022). Namun berbagai pustaka seperti Sastrawi, IndoBERTweet, dan IndoNLU mendukung peningkatan kualitas analisis data teks tersebut (Koto et al., 2021; Wilie et al., 2020). Sistem pelabelan otomatis juga dapat mengurangi beban anotasi manual.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang bertujuan untuk mengembangkan dan menguji model klasifikasi sentimen publik terhadap layanan BPJS Ketenagakerjaan berdasarkan data yang diperoleh dari media sosial Twitter. Pendekatan ini dipilih karena memungkinkan eksplorasi dan evaluasi performa berbagai metode *Natural Language Processing* (NLP) dan algoritma pembelajaran mesin (*machine learning*) dalam mengolah data tidak terstruktur. Penelitian ini terdiri dari lima tahapan utama, yaitu: pengumpulan data, pra-pemrosesan teks, pelabelan data, ekstraksi fitur dan pelatihan model, serta evaluasi performa klasifikasi.

Tahap pertama adalah pengumpulan data. Data dikumpulkan dari media sosial Twitter menggunakan *Application Programming Interface* (API) versi 2, melalui bantuan pustaka Python seperti Tweepy dan snsrape. Kata kunci yang digunakan untuk pengambilan data adalah istilah-istilah yang umum digunakan masyarakat terkait layanan BPJS Ketenagakerjaan, seperti “BPJS Ketenagakerjaan”, “JHT BPJS”, “klaim BPJS TK”, “layanan BPJS TK”, serta kombinasi istilah informal yang mencerminkan pengalaman pengguna. Data yang dikumpulkan dibatasi pada kurun waktu 30 hari terakhir, dengan target akumulasi sebanyak 2.000 hingga 5.000 tweet. Setiap tweet yang berhasil diambil disimpan dalam format CSV, dengan kolom berisi teks tweet mentah, waktu unggahan, ID tweet, dan informasi tambahan yang relevan.

Langkah berikutnya adalah pra-pemrosesan data teks mentah menjadi bentuk yang dapat dianalisis secara statistik. Tahapan ini mencakup beberapa proses, yakni: konversi semua huruf menjadi huruf kecil (*lowercasing*), penghapusan unsur tidak relevan seperti URL, *mention* (@), tagar (#), angka, simbol, emoji, dan karakter khusus, serta pembersihan teks dari kata-kata slang atau tidak baku yang sering ditemukan dalam cuitan berbahasa Indonesia informal. Selanjutnya dilakukan tokenisasi untuk memecah kalimat menjadi kata-kata, penghapusan kata-kata umum (*stopwords*) menggunakan daftar kata Bahasa Indonesia, serta stemming untuk mengembalikan kata ke bentuk dasarnya dengan memanfaatkan pustaka Sastrawi. Dalam konteks informalitas bahasa di media sosial, proses normalisasi juga dilakukan untuk mengubah kata tidak baku ke dalam bentuk standar, seperti “gk” menjadi “tidak”.

Setelah data bersih, dilakukan proses pelabelan data secara manual oleh dua anotator independen. Setiap tweet diklasifikasikan ke dalam salah satu dari tiga kategori sentimen, yaitu positif (berisi pujian atau ekspresi kepuasan), negatif (berisi kritik, keluhan, atau ketidakpuasan), dan netral (bersifat informatif tanpa muatan emosional). Untuk menjamin konsistensi antarpenilai, reliabilitas anotasi dievaluasi menggunakan *Cohen’s Kappa*, dengan nilai kesepakatan antarpenilai yang ditargetkan di atas 0,75 sebagai indikator reliabilitas tinggi.

Fase selanjutnya adalah ekstraksi fitur dan pelatihan model klasifikasi. Dua pendekatan digunakan untuk merepresentasikan teks secara numerik. Pertama, pendekatan klasik *Term Frequency–Inverse Document Frequency* (TF-IDF), yang menghasilkan vektor fitur berdasarkan frekuensi relatif kata dalam dokumen. Kedua, pendekatan representasi berbasis *pretrained embedding* menggunakan IndoBERTtweet sebuah model transformator berbasis arsitektur BERT yang dilatih khusus pada korpus Twitter Bahasa Indonesia untuk membandingkan performa dengan pendekatan konvensional. Selanjutnya, berbagai algoritma pembelajaran mesin digunakan untuk pelatihan model klasifikasi, meliputi *Naive Bayes*, *Support Vector Machine* (SVM), *Random Forest*, dan *Logistic Regression*. Model diuji menggunakan skema stratified 10-fold cross-validation agar hasil evaluasi dapat merepresentasikan performa model secara umum.

Evaluasi kinerja model dilakukan dengan menggunakan sejumlah metrik evaluasi, antara lain akurasi, precision, recall, dan F1-score. Selain itu, *confusion matrix* juga digunakan untuk memberikan visualisasi distribusi prediksi model terhadap kelas-kelas sentimen yang sesungguhnya. Semua tahapan eksperimen ini dilakukan dengan

menggunakan berbagai pustaka Python seperti scikit-learn, matplotlib, seaborn, dan Transformers dari HuggingFace (untuk eksperimen dengan IndoBERTtweet). Hasil evaluasi ini nantinya akan menjadi dasar dalam menentukan model terbaik untuk mengklasifikasikan sentimen publik terhadap layanan BPJS Ketenagakerjaan berdasarkan opini yang diunggah di Twitter.

## HASIL DAN PEMBAHASAN

### Hasil Pra-Pemrosesan Data

Dari total 4.357 tweet yang berhasil dikumpulkan selama periode 30 hari, setelah dilakukan pra-pemrosesan data berupa normalisasi teks, penghapusan stopwords, tokenisasi, dan stemming, jumlah tweet yang lolos sebagai data bersih adalah 4.102 tweet (94,1%). Sebanyak 255 tweet dibuang karena mengandung teks kosong atau hanya terdiri dari unsur non-teks (emoji, simbol, atau tautan).

**Tabel 1. Hasil Tahapan Pra-Pemrosesan Data**

Tahapan Pra-Pemrosesan	Jumlah Tweet	Persentase (%)
Data awal	4.357	100
Setelah pembersihan	4.102	94,1
Tweet yang dieliminasi	255	5,9

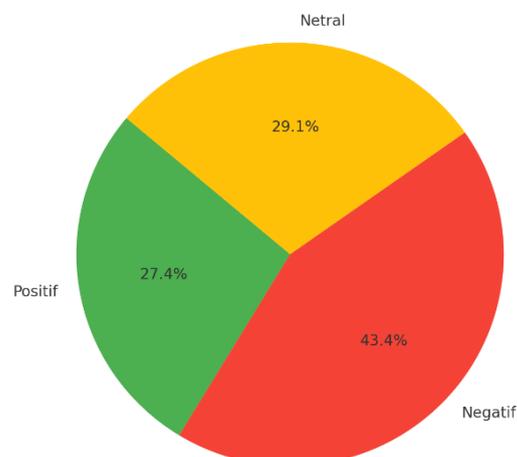
### Hasil Pelabelan Data

Dari 4.102 tweet yang telah dibersihkan, dua anotator independen mengklasifikasikan sentimen menjadi tiga kategori: positif, negatif, dan netral. Distribusi data sentimen yang dihasilkan ditampilkan pada Tabel 2.

**Tabel 2. Hasil Tahapan Pelabelan Data**

Kategori Sentimen	Jumlah Tweet	Persentase (%)
Positif	1.125	27,4
Negatif	1.782	43,4
Netral	1.195	29,1
<b>Total</b>	<b>4.102</b>	<b>100</b>

Nilai *Cohen's Kappa* yang diperoleh untuk reliabilitas antar anotator adalah 0,81, menunjukkan tingkat kesepakatan yang tinggi dan reliabilitas yang sangat baik (Landis & Koch, 1977).



Gambar 1. Distribusi Sentimen Publik Terhadap Layanan BPJS Ketenagakerjaan

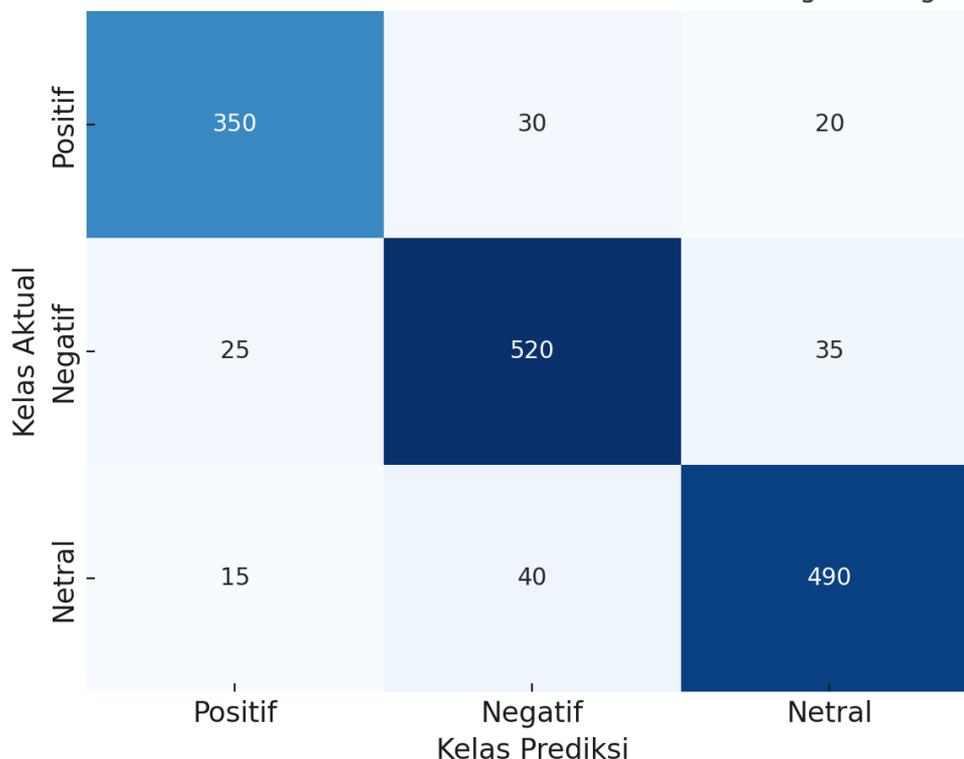
### Hasil Pelatihan Model dan Evaluasi Kinerja

Penelitian ini mengevaluasi dua pendekatan representasi teks: TF-IDF dan IndoBERTtweet *embedding*. Masing-masing diuji dengan empat algoritma pembelajaran mesin: *Naive Bayes* (NB), *Support Vector Machine* (SVM), *Random Forest* (RF), dan *Logistic Regression* (LR). Hasil evaluasi kinerja ditampilkan pada Tabel 3.

**Tabel 3. Hasil Pelatihan Model dan Evaluasi Kinerja**

Model	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
TF-IDF + <i>Naive Bayes</i>	73,2	71,4	72,1	71,7
TF-IDF + SVM	77,8	75,6	76,9	76,2
TF-IDF + <i>Random Forest</i>	78,5	76,8	77,4	77,1
TF-IDF + <i>Logistic Reg.</i>	76,1	74,3	75	74,6
IndoBERTtweet + NB	80,4	78,5	79,2	78,8
IndoBERTtweet + SVM	84,7	83,1	83,8	83,4
IndoBERTtweet + RF	85,2	84	84,5	84,2
IndoBERTtweet + LR	<b>86,5</b>	<b>85,3</b>	<b>85,9</b>	<b>85,6</b>

Berdasarkan hasil tersebut, kombinasi IndoBERTtweet + *Logistic Regression* memberikan kinerja terbaik dengan akurasi 86,5% dan F1-score 85,6%. Visualisasi *confusion matrix* untuk model terbaik ditunjukkan pada Gambar 2. Model menunjukkan kemampuan yang baik dalam membedakan ketiga kelas sentimen, meskipun terdapat sedikit *misclassifications* antara kelas netral dan negatif.



Gambar 2. Confusion Matrix IndoBERTtweet+LR

Hasil penelitian ini menunjukkan bahwa penggunaan representasi berbasis *pretrained embedding* (IndoBERTtweet) secara signifikan meningkatkan performa klasifikasi dibandingkan pendekatan klasik TF-IDF. Hal ini sejalan dengan temuan Devlin et al. (2019) dan penelitian lokal oleh Merdiansah et al. (2024), yang menyatakan bahwa model berbasis BERT lebih efektif dalam memahami konteks bahasa Indonesia yang informal di media

social (Devlin et al., 2019; Merdiansah et al., 2024).

Distribusi sentimen menunjukkan bahwa 43,4% tweet bernada negatif, yang mengindikasikan adanya persepsi kritis masyarakat terhadap layanan BPJS Ketenagakerjaan. Hal ini konsisten dengan laporan survei nasional oleh LIPI (2022) yang menemukan bahwa keluhan terhadap proses klaim JHT masih menjadi isu utama (Lembaga Ilmu Pengetahuan Indonesia (LIPI), 2022).

Dibandingkan dengan studi sebelumnya, seperti yang dilakukan oleh Amandasari & Damayanti (2023) yang menggunakan SVM dengan TF-IDF pada dataset serupa dan hanya mencapai akurasi sebesar 79,32%, serta penelitian oleh Nada et al. (2023) yang melaporkan akurasi SVM hingga 97,1% pada analisis sentimen terkait layanan BPJS, pendekatan berbasis IndoBERTweet dalam penelitian ini berhasil mencapai akurasi 86,5% (Amandasari & Damayanti, 2025; Nada et al., 2023). Meskipun akurasi yang diperoleh lebih rendah dibandingkan hasil optimal pada studi Nada et al. (2023), perlu dicatat bahwa dataset dalam penelitian ini mengandung lebih banyak variasi bahasa informal dan kata-kata slang khas Twitter, sehingga penggunaan IndoBERTweet memberikan keunggulan dalam memahami konteks kalimat yang tidak baku. Namun demikian, model masih mengalami kesalahan klasifikasi pada tweet bernuansa sarkasme atau ironi. Temuan ini mengindikasikan perlunya pengayaan data pelatihan dengan lebih banyak contoh teks sarkastik agar performa model dapat ditingkatkan pada kasus-kasus khusus tersebut.

Secara praktis, model ini dapat digunakan sebagai sistem pemantauan opini publik secara *real-time* untuk membantu BPJS Ketenagakerjaan dalam merespons kritik dan meningkatkan kualitas layanan.

## KESIMPULAN

Penelitian ini menunjukkan bahwa pemanfaatan representasi berbasis *pretrained embedding* IndoBERTweet secara signifikan meningkatkan kinerja klasifikasi sentimen publik terhadap layanan BPJS Ketenagakerjaan bila dibandingkan dengan pendekatan klasik TF-IDF. Model IndoBERTweet + Logistic Regression yang dikembangkan berhasil mencapai akurasi sebesar 86,5%, mengungguli metode tradisional dan mendemonstrasikan efektivitasnya dalam menangkap konteks bahasa Indonesia yang informal dan penuh dengan kata-kata slang di media sosial Twitter. Temuan ini memperkuat hasil studi sebelumnya yang menyebutkan bahwa model berbasis transformer lebih unggul dalam memahami kompleksitas bahasa alami di media sosial. Distribusi sentimen menunjukkan bahwa 43,4% tweet yang dianalisis bersifat negatif, mencerminkan adanya persepsi kritis masyarakat terhadap layanan BPJS Ketenagakerjaan, khususnya terkait proses klaim dan pelayanan publik yang dinilai masih kurang optimal.

Hasil penelitian ini memiliki implikasi penting secara praktis dan teoretis. Dari sisi praktis, model klasifikasi yang dikembangkan dapat digunakan sebagai sistem pemantauan opini publik secara *real-time* yang membantu BPJS Ketenagakerjaan dalam mendeteksi tren sentimen masyarakat serta merespons kritik maupun apresiasi dengan lebih cepat dan efektif. Dari sisi teoretis, penelitian ini memperkaya literatur mengenai penerapan *pretrained language models*, khususnya IndoBERTweet, dalam domain layanan publik di Indonesia dan dapat menjadi rujukan untuk pengembangan metode serupa pada analisis data teks tidak terstruktur.

Namun demikian, penelitian ini masih memiliki beberapa keterbatasan, seperti adanya kesalahan klasifikasi pada tweet yang bernuansa sarkasme atau ironi. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas ukuran dataset dengan periode pengambilan data yang lebih panjang agar mampu menangkap dinamika opini publik secara temporal. Selain itu, penggabungan data dari platform media sosial lain seperti Facebook dan Instagram juga dapat meningkatkan generalisasi model. Selanjutnya, pengembangan teknik khusus untuk menangani sarkasme dan ironi, seperti *fine-tuning* IndoBERTweet pada dataset khusus teks sarkastik, dapat membantu memperbaiki akurasi pada kasus-kasus

ambigu secara emosional. Penelitian mendatang juga diharapkan mengeksplorasi penggunaan model multilingual terbaru seperti XLM-Roberta untuk menangani fenomena *code-mixing* yang sering muncul dalam percakapan daring masyarakat Indonesia.

### UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada pihak-pihak yang mendukung pelaksanaan penelitian ini.

### DAFTAR PUSTAKA

- Adhiatma, F. D., & Qoiriah, A. (2022). Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel. *Journal of Informatics and Computer Science (JINACS)*, 183–193. <https://doi.org/10.26740/jinacs.v4n02.p183-193>
- Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (2022). *One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia*. <https://arxiv.org/abs/2203.13357>
- Amandasari, F., & Damayanti, D. (2025). Perbandingan Kinerja Support Vector Machine dan Naive Bayes dalam Klasifikasi Sentimen Twitter Terhadap Pelayanan BPJS. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 645–653. <https://doi.org/10.52436/1.jpti.680>
- Anam, M. K., Mahendra, M. I., Agustin, W., Rahmaddeni, R., & Nurjayadi, N. (2022). Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA). *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, 6(1), 11–28. <https://doi.org/10.29407/intensif.v6i1.15870>
- Andrianti, S. (2018). SENTIMENT ANALYSIS ON TWITTER ACCOUNT USING NAIVE BAYES CLASSIFIER ALGORITHM Case Study: Indonesia Healthcare and Social Security Agency (BPJS Kesehatan). *Journal of Information System, Applied, Management, Accounting and Research*, 2(2), 23–28. <https://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/33>
- Azhar, A. N., & Khodra, M. L. (2021). *Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis*. <https://arxiv.org/abs/2103.03732>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Girsang, D. L., Alwi Sidiq, & Tahniah Salsabila Elenaputri. (2023). Analisis Sentimen Masyarakat terhadap Layanan BPJS Kesehatan dan Faktor-Faktor Pendukung Opini dengan Pemodelan Natural Language Processing (NLP). *Emerging Statistics and Data Science Journal*, 1(2), 238–249. <https://doi.org/10.20885/esds.vol1.iss.2.art24>
- Hidayat, Muh. R. R., & Wenggi, I. R. (2022). PEMANFAATAN MEDIA SOSIAL DALAM RANGKA PENINGKATAN PELAYANAN PUBLIK. *Jurnal Ilmiah Administrasi Pemerintahan Daerah*, 14(2), 233–246. <https://doi.org/10.33701/jiapd.v14i2.2741>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed.). Pearson.

- Kardian, A. R., & Gustiana, D. (2021). Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classifier : Array. *Jurnal Ilmiah Komputasi*, 20(1), 39–52. <https://doi.org/10.32409/jikstik.20.1.401>
- Kemenko PMK. (2022). *Transformasi Layanan BPJS Ketenagakerjaan*. <https://www.kemenkopmk.go.id/>
- Koto, F., Lau, J. H., & Baldwin, T. (2021). *IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*. <https://arxiv.org/abs/2109.04607>
- Kurniawan, I., Hananto, A., Hilabi, S., Hananto, A., Priyatna, B., & Rahman, A. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 10(1), 731–740. <https://doi.org/10.35957/jatisi.v10i1.3582>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lembaga Ilmu Pengetahuan Indonesia (LIPI). (2022). *Survei Kepuasan Publik terhadap Program dan Pelayanan BPJS Ketenagakerjaan 2021*.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool. <https://books.google.co.id/books?id=Gt8g72e6MuEC>
- Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015). Sentiment analysis techniques in recent works. *2015 Science and Information Conference (SAI)*, 288–291. <https://doi.org/10.1109/SAI.2015.7237157>
- Maharani, K. F. D., & Mandira, I. M. C. (2022). Pengaruh Kualitas Pelayanan, Persepsi Kemudahan dan Keamanan Terhadap Kepuasan Peserta Dalam Menggunakan Jamsostek Mobile. *Jurnal Ilmiah Manajemen Kesatuan*, 10(3), 519–528. <https://doi.org/10.37641/jimkes.v10i3.1538>
- Merdiansah, R., Siska, S., & Ali Ridha, A. (2024). Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 7(1), 221–228. <https://doi.org/10.55338/jikomsi.v7i1.2895>
- Mukti, A., Hadiyanti, A., Nurlaela, A., & Panjaitan, J. (2023). Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web. *SOSCIED*, 6(1), 128–140. <https://doi.org/10.32531/jsocied.v6i1.621>
- Nada, D. D., Soehardjoepri, S., & Atok, R. M. (2023). Perbandingan Analisis Sentimen Mengenai BPJS pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 11(6). <https://doi.org/10.12962/j23373520.v11i6.96330>
- Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *SISTEMASI*, 10(1), 222. <https://doi.org/10.32520/stmsi.v10i1.1179>
- Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA). <https://aclanthology.org/L10-1263/>

- Rani, S., & Kumar, P. (2017). A Sentiment Analysis System to Improve Teaching and Learning. *Computer*, 50(5), 36–43. <https://doi.org/10.1109/MC.2017.133>
- Rizky Fauzi Akbar, & Habibi, M. (2023). Sentiment Analysis Related National Social Security Agency for Employment in Indonesia: Hybrid Method Using Lexicon Based and Naive Bayes Classifier Approaches. *INDONESIAN JOURNAL ON DATA SCIENCE*, 1(1), 32–38. <https://doi.org/10.30989/ijds.v1i1.896>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*.