

ANALISIS SENTIMEN DAN IDENTIFIKASI ISU KRITIS LAYANAN BPJS KESEHATAN PADA MEDIA SOSIAL X MENGGUNAKAN INDOBERTWEET DAN LDA

Nur Insan Subekti^{1*}, Asriyanik², Didik Indrayana³

^{1, 2, 3} Teknik Informatika, Universitas Muhammadiyah Sukabumi

Keywords:

Analisis Sentimen;
IndoBERTweet;
Latent Dirichlet Allocation;
BPJS Kesehatan.

Correspondent Email:

nurinsan058@ummi.ac.id

Abstrak. Perkembangan media sosial menjadikan platform digital sebagai sumber data yang dapat digunakan untuk memahami persepsi masyarakat terhadap layanan publik. Media sosial X menyediakan berbagai opini pengguna dalam bentuk teks, namun karakteristik data yang tidak terstruktur, penggunaan bahasa informal, serta jumlah data yang besar menyebabkan analisis secara manual menjadi tidak efektif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat dan mengidentifikasi isu kritis layanan BPJS Kesehatan berdasarkan opini pengguna pada media sosial X menggunakan IndoBERTweet dan Latent Dirichlet Allocation (LDA). Penelitian menggunakan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM) yang meliputi tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Sebanyak 5.336 cuitan dikumpulkan dari media sosial X, kemudian 1.000 data digunakan sebagai dataset anotasi manual untuk proses continued fine-tuning model IndoBERTweet. Model hasil fine-tuning dievaluasi menggunakan accuracy, precision, recall, dan F1-score sebelum diterapkan untuk klasifikasi keseluruhan dataset. Data dengan sentimen negatif selanjutnya dianalisis menggunakan LDA untuk memperoleh kelompok topik isu kritis. Hasil penelitian menunjukkan bahwa kombinasi IndoBERTweet dan LDA mampu mengubah data opini masyarakat menjadi informasi terstruktur berupa kategori sentimen dan kelompok isu layanan BPJS Kesehatan. Penelitian ini memberikan pendekatan berbasis NLP untuk mendukung evaluasi layanan publik berdasarkan data media sosial.



Copyright © 2026 The Author(s),
Published by Jurnal Informatika dan
Teknik Elektro Terapan. This is an
open-access article distributed under
the terms of the Creative Commons
Attribution-NonCommercial 4.0
International License (CC BY-NC
4.0).

Abstract. The development of social media has made digital platforms an important source for understanding public perceptions of public services. Social media X provides user opinions in textual form; however, the unstructured characteristics, informal language usage, and large data volume make manual analysis inefficient. This study aims to analyze public sentiment and identify critical issues regarding BPJS Kesehatan services based on user opinions on social media X using IndoBERTweet and Latent Dirichlet Allocation (LDA). The research applies the Cross Industry Standard Process for Data Mining (CRISP-DM) framework consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. A total of 5,336 tweets were collected from social media X, where 1,000 manually annotated data were used for continued fine-tuning of the IndoBERTweet model. The fine-tuned model was evaluated using accuracy, precision, recall, and F1-score before being applied to classify the entire dataset. Negative sentiment data were further analyzed using LDA to identify critical issue topics. The results demonstrate that the combination of IndoBERTweet and LDA can transform public opinions into structured information consisting of sentiment categories and issue clusters related to

BPJS Kesehatan services. This research provides an NLP-based approach to support public service evaluation using social media data.

1. PENDAHULUAN

Transformasi teknologi informasi telah mengubah pola komunikasi masyarakat dari ruang publik konvensional menuju ruang digital. Media sosial saat ini tidak hanya digunakan sebagai sarana komunikasi, tetapi juga menjadi sumber data yang menggambarkan opini, pengalaman, serta respons masyarakat terhadap berbagai isu publik. Salah satu platform yang banyak digunakan untuk menyampaikan opini secara langsung adalah X, yang memiliki karakteristik real-time dan terbuka sehingga dapat dimanfaatkan sebagai sumber data dalam memahami persepsi masyarakat terhadap suatu layanan [1].

BPJS Kesehatan merupakan salah satu layanan publik strategis dengan cakupan peserta yang sangat besar di Indonesia. Peningkatan jumlah peserta menyebabkan kualitas pelayanan menjadi aspek penting karena layanan yang diberikan mencakup berbagai proses, mulai dari administrasi kepesertaan, pelayanan fasilitas kesehatan, sistem rujukan, hingga layanan digital melalui aplikasi Mobile JKN [2]. Namun, kompleksitas layanan tersebut masih menimbulkan berbagai pengalaman pengguna, baik berupa apresiasi maupun keluhan terkait proses pelayanan, administrasi, dan akses layanan kesehatan [3].

Besarnya jumlah opini masyarakat yang disampaikan melalui media sosial menyebabkan proses identifikasi permasalahan secara manual menjadi kurang efektif. Data teks pada media sosial memiliki karakteristik tidak terstruktur, penggunaan bahasa informal, singkatan, serta variasi penulisan yang menjadi tantangan dalam pengolahan bahasa alami (Natural Language Processing/NLP). Oleh karena itu, diperlukan pendekatan berbasis kecerdasan buatan yang mampu mengolah data teks secara otomatis untuk memperoleh informasi mengenai kecenderungan sentimen masyarakat [4], [5].

Analisis sentimen merupakan salah satu pendekatan NLP yang digunakan untuk mengklasifikasikan opini berdasarkan polaritas tertentu, seperti positif dan negatif. Namun,

analisis sentimen hanya memberikan informasi mengenai kecenderungan opini tanpa menjelaskan aspek layanan yang menjadi penyebab munculnya sentimen tersebut. Oleh karena itu, diperlukan pendekatan tambahan berupa pemodelan topik untuk mengidentifikasi isu atau pola permasalahan yang terdapat dalam kumpulan opini masyarakat [6], [7].

Penelitian sebelumnya menunjukkan bahwa model berbasis Transformer memiliki kemampuan yang baik dalam memahami konteks bahasa dibandingkan metode berbasis fitur tradisional [8]. IndoBERTweet merupakan model bahasa berbasis Transformer yang dikembangkan menggunakan korpus media sosial berbahasa Indonesia sehingga mampu memahami karakteristik bahasa informal pada platform X [9]. Melalui proses continued fine-tuning, model dapat disesuaikan dengan karakteristik domain tertentu, seperti opini masyarakat mengenai layanan BPJS Kesehatan [10].

Selain klasifikasi sentimen, identifikasi isu kritis dapat dilakukan menggunakan Latent Dirichlet Allocation (LDA), yaitu algoritma pemodelan topik yang mampu menemukan pola topik tersembunyi pada kumpulan dokumen teks. Integrasi antara analisis sentimen dan pemodelan topik memungkinkan penelitian tidak hanya mengetahui bagaimana persepsi masyarakat, tetapi juga memahami aspek layanan yang menjadi perhatian utama pengguna.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan BPJS Kesehatan pada media sosial X menggunakan model IndoBERTweet hasil continued fine-tuning serta mengidentifikasi isu kritis berdasarkan sentimen negatif menggunakan metode LDA.

2. TINJAUAN PUSTAKA

1. Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang ilmu kecerdasan buatan yang berfokus pada kemampuan komputer dalam memahami, mengolah, dan menghasilkan bahasa manusia secara

komputasional. NLP memungkinkan data teks yang bersifat tidak terstruktur diubah menjadi informasi yang dapat dianalisis melalui berbagai pendekatan, seperti klasifikasi teks, ekstraksi informasi, dan pemodelan topik [11].

Dalam penelitian berbasis media sosial, NLP memiliki tantangan tersendiri karena data yang digunakan umumnya memiliki karakteristik berupa bahasa informal, penggunaan singkatan, kesalahan penulisan, serta variasi penggunaan kata. Oleh karena itu, tahapan preprocessing dan pemilihan model representasi teks menjadi faktor penting dalam menghasilkan analisis yang akurat.

2. Analisis Sentimen

Analisis sentimen merupakan penerapan NLP yang bertujuan untuk mengidentifikasi kecenderungan opini atau emosi yang terkandung dalam suatu teks. Pendekatan ini banyak digunakan untuk memahami persepsi pengguna terhadap suatu produk, layanan, maupun kebijakan berdasarkan data tekstual dari media sosial [5].

Pada penelitian ini, analisis sentimen digunakan untuk mengelompokkan opini masyarakat terhadap layanan BPJS Kesehatan ke dalam dua kategori, yaitu sentimen positif dan negatif. Hasil klasifikasi sentimen digunakan sebagai dasar untuk mengetahui kecenderungan persepsi masyarakat serta memilih data dengan sentimen negatif untuk tahap identifikasi isu kritis menggunakan pemodelan topik.

3. IndoBERTweet dan Continued Fine-tuning

Transformer merupakan arsitektur deep learning yang banyak digunakan dalam pemrosesan bahasa alami karena mampu memahami hubungan kontekstual antar kata melalui mekanisme self-attention. Salah satu model berbasis Transformer yang dikembangkan untuk bahasa Indonesia adalah IndoBERTweet.

IndoBERTweet merupakan model bahasa berbasis BERT yang dilatih menggunakan kumpulan data media sosial berbahasa Indonesia. Model ini dirancang untuk memahami karakteristik bahasa pada platform X, termasuk penggunaan slang, singkatan, dan variasi bahasa informal yang sering muncul pada cuitan pengguna [9].

Meskipun model pre-trained telah memiliki kemampuan memahami struktur bahasa secara umum, adaptasi terhadap domain tertentu diperlukan agar model dapat memahami konteks yang lebih spesifik. Pada penelitian ini dilakukan continued fine-tuning menggunakan data cuitan terkait layanan BPJS Kesehatan yang telah diberi label secara manual. Pendekatan tersebut bertujuan untuk menyesuaikan representasi bahasa IndoBERTweet terhadap karakteristik opini masyarakat mengenai layanan kesehatan [10].

Beberapa penelitian menunjukkan bahwa model berbasis Transformer seperti IndoBERTweet memiliki performa yang baik dalam klasifikasi teks berbahasa Indonesia dibandingkan metode berbasis machine learning tradisional [12].

4. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan algoritma unsupervised learning yang digunakan untuk menemukan pola topik tersembunyi dalam kumpulan dokumen teks. LDA mengasumsikan bahwa setiap dokumen terdiri dari beberapa distribusi topik, sedangkan setiap topik terdiri dari kumpulan kata yang memiliki hubungan probabilistik tertentu [7].

Dalam penelitian ini, LDA digunakan untuk mengidentifikasi isu kritis pada cuitan bersentimen negatif terkait layanan BPJS Kesehatan. Hasil pemodelan topik digunakan untuk mengetahui aspek layanan yang paling sering menjadi perhatian masyarakat [7].

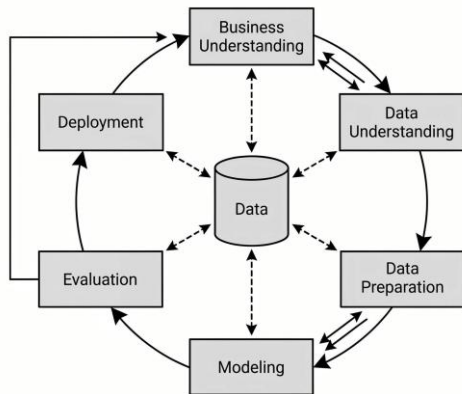
Jumlah topik optimal pada LDA ditentukan menggunakan nilai coherence score. Nilai coherence yang lebih tinggi menunjukkan bahwa kata-kata dalam suatu topik memiliki keterkaitan semantik yang lebih baik [13], [14].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja penelitian. CRISP-DM dipilih karena mampu memberikan tahapan sistematis dalam proses analisis data, mulai dari pemahaman masalah hingga implementasi hasil analisis [15].

Tahapan penelitian terdiri dari:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment



1. Business Understanding

Tahap business understanding dilakukan untuk memahami permasalahan layanan BPJS Kesehatan berdasarkan opini masyarakat pada media sosial X. Permasalahan utama yang dikaji adalah bagaimana mengidentifikasi kecenderungan sentimen masyarakat serta menemukan isu kritis yang menyebabkan munculnya opini negatif terhadap layanan BPJS Kesehatan.

2. Data Understanding

Data penelitian diperoleh melalui proses web scraping terhadap platform X menggunakan kata kunci yang berkaitan dengan layanan BPJS Kesehatan.

Dataset awal terdiri dari 5.336 cuitan yang dikumpulkan pada periode Januari 2024 hingga Juni 2026.

Atribut data yang digunakan meliputi:

Atribut	Keterangan
tanggal	Waktu publikasi cuitan
username	Identitas akun pengguna
tweet	Isi teks cuitan

Data yang digunakan hanya berupa teks publik tanpa menggunakan informasi pribadi

pengguna.

3. Data Preparation

Tahap persiapan data dilakukan untuk meningkatkan kualitas data sebelum proses pemodelan. Tahapan preprocessing meliputi: Cleaning

Menghapus elemen yang tidak diperlukan seperti URL, mention, simbol, dan karakter khusus.

1) Case Folding

Mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan bentuk kata.

2) Normalisasi Slang

Mengubah kata tidak baku atau singkatan menjadi bentuk kata yang lebih standar.

3) Penghapusan Noise

Menghapus elemen yang tidak memberikan informasi penting terhadap analisis.

4) Stopword Removal

Menghapus kata umum yang tidak memiliki kontribusi besar terhadap proses analisis.

5) Tokenisasi

Mengubah teks menjadi token menggunakan tokenizer IndoBERTweet.

4. Data Labeling dan Pembagian Dataset

Sebanyak 1.000 data cuitan dilakukan anotasi manual untuk menghasilkan dataset berlabel.

Label sentimen dibagi menjadi:

0 = Negatif

1 = Positif

Dataset kemudian dibagi menggunakan metode stratified split menjadi:

Dataset	Jumlah Data
Training	800
Validation	100
Testing	100

Pembagian dilakukan secara stratifikasi untuk menjaga keseimbangan distribusi kelas sentimen.

5. Modeling

- 1) Continued Fine-tuning IndoBERTweet Model IndoBERTweet digunakan sebagai model dasar untuk melakukan klasifikasi sentimen. Proses continued fine-tuning dilakukan menggunakan dataset hasil anotasi manual.

Parameter pelatihan model:

Parameter	Nilai
Learning rate	2e-5
Batch size	16
Epoch	5
Weight decay	0.01

Model hasil fine-tuning kemudian digunakan untuk melakukan prediksi sentimen terhadap data yang belum memiliki label.

2) Pemodelan Topik LDA

Data dengan prediksi sentimen negatif digunakan sebagai masukan untuk proses pemodelan topik menggunakan LDA.

Tahapan LDA meliputi:

- a) Seleksi data negatif
- b) Preprocessing khusus pemodelan topik
- c) Pembentukan dictionary dan corpus
- d) Pengujian beberapa jumlah topik
- e) Pemilihan jumlah topik berdasarkan coherence score
- f) Interpretasi topik

6. Evaluation

Evaluasi model klasifikasi dilakukan menggunakan:

- 1) Accuracy
- 2) Precision
- 3) Recall
- 4) F1-score

Sedangkan evaluasi pemodelan topik dilakukan menggunakan coherence score untuk mengetahui kualitas keterkaitan antar kata dalam setiap topik.

7. Deployment

Hasil penelitian diimplementasikan dalam aplikasi berbasis Streamlit yang menyediakan fitur:

- 1) Analisis sentimen teks tunggal
- 2) Analisis sentimen berbasis file CSV
- 3) Identifikasi isu kritis berdasarkan hasil pemodelan topik

4. HASIL DAN PEMBAHASAN

1. Hasil Pengumpulan dan Persiapan Data

Penelitian ini menggunakan data cuitan dari media sosial X yang berkaitan dengan layanan BPJS Kesehatan. Proses pengumpulan data menghasilkan sebanyak 5.336 cuitan yang kemudian melalui tahapan seleksi dan preprocessing untuk menghilangkan data yang tidak relevan serta meningkatkan kualitas teks sebelum dilakukan analisis.

Tahapan preprocessing dilakukan untuk menyesuaikan karakteristik data media sosial yang umumnya mengandung noise seperti URL, mention, simbol, penggunaan huruf tidak konsisten, serta istilah tidak baku. Tahapan preprocessing yang diterapkan meliputi cleaning, case folding, penghapusan noise, normalisasi slang, stopword removal, dan tokenisasi menggunakan tokenizer IndoBERTweet.

Contoh hasil preprocessing dapat dilihat pada Tabel 1.

Tahap	Teks
Sebelum preprocessing	"@BPJSKesehatan min kok aplikasi Mobile JKN error terus ya? susah daftar antrean 🤔 https://xxx "
Setelah preprocessing	"bpjs kesehatan aplikasi mobile jkn error daftar antrean"

Hasil preprocessing menunjukkan bahwa teks menjadi lebih terstruktur sehingga dapat digunakan sebagai masukan model klasifikasi berbasis Transformer.

2. Hasil Anotasi Manual dan Pembagian Dataset

Sebanyak 1.000 data cuitan digunakan sebagai dataset berlabel untuk proses continued fine-tuning IndoBERTweet. Proses anotasi dilakukan secara manual dengan dua kategori sentimen, yaitu positif dan negatif.

Distribusi dataset hasil anotasi kemudian dibagi menggunakan metode stratified split untuk menjaga keseimbangan distribusi kelas.

Tabel 2. Distribusi Dataset Continued Fine-tuning

Dataset	Jumlah Data	Negatif	Positif
Training	800	401	399
Validation	100	50	50
Testing	100	50	50

Berdasarkan pembagian tersebut, dataset memiliki distribusi kelas yang relatif seimbang sehingga model dapat mempelajari pola kedua kategori sentimen secara proporsional.

3. Hasil Continued Fine-tuning Model IndoBERTweet

Proses continued fine-tuning dilakukan untuk menyesuaikan kemampuan model IndoBERTweet terhadap karakteristik opini masyarakat mengenai layanan BPJS Kesehatan. Model dilatih menggunakan dataset training dengan parameter pembelajaran sebagai berikut:

Parameter	Nilai
Learning rate	2e-5
Batch size	16
Epoch	5
Weight decay	0,01
Optimizer	AdamW

Selama proses pelatihan, model melakukan pembaruan parameter berdasarkan perbedaan antara prediksi model dan label aktual. Data validation digunakan untuk memantau kemampuan model selama proses pelatihan, sedangkan data testing digunakan untuk mengukur kemampuan generalisasi model.

4. Evaluasi Model Klasifikasi Sentimen

Evaluasi model dilakukan menggunakan data testing sebanyak 100 data yang terdiri dari 50 data negatif dan 50 data positif.

Kelas	Precision	Recall	F1-score
Negatif	0,9375	0,9375	0,9375
Positif	0,9423	0,9423	0,9423
Accuracy			0,94

Berdasarkan hasil evaluasi, model memperoleh nilai accuracy sebesar 94%. Nilai precision, recall, dan F1-score yang relatif seimbang menunjukkan bahwa model mampu mengenali kedua kategori sentimen secara baik.

Confusion matrix hasil pengujian menunjukkan:

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	45	3
Aktual Positif	3	49

Berdasarkan confusion matrix tersebut, terdapat 94 data yang berhasil diklasifikasikan dengan benar dari total 100 data pengujian. Hasil tersebut menunjukkan bahwa pendekatan continued fine-tuning mampu meningkatkan kemampuan IndoBERTweet dalam memahami konteks bahasa pada domain layanan BPJS Kesehatan. Hal ini disebabkan karena model tidak hanya menggunakan pengetahuan bahasa

umum dari tahap pre-training, tetapi juga memperoleh adaptasi terhadap pola bahasa, istilah layanan kesehatan, serta ekspresi opini masyarakat yang terdapat pada dataset penelitian.

5. Hasil Inferensi Dataset dan Distribusi Sentimen

Model IndoBERTweet hasil continued fine-tuning selanjutnya diterapkan untuk melakukan prediksi terhadap data yang belum memiliki label. Proses inferensi dilakukan terhadap keseluruhan dataset penelitian untuk memperoleh gambaran umum persepsi masyarakat mengenai layanan BPJS Kesehatan. Hasil klasifikasi menunjukkan distribusi sentimen sebagai berikut:

Sentimen	Jumlah Data	Persentase
Positif	2.673	50,09%
Negatif	2.663	49,91%
Total	5.336	100%

Hasil tersebut menunjukkan bahwa distribusi sentimen masyarakat relatif seimbang. Sentimen positif memiliki selisih yang sangat kecil dibandingkan sentimen negatif, yaitu sebesar 10 data.

Dominasi sentimen positif menunjukkan bahwa sebagian masyarakat memberikan respons yang baik terhadap layanan BPJS Kesehatan. Namun, jumlah sentimen negatif yang hampir setara menunjukkan masih terdapat berbagai permasalahan layanan yang dirasakan pengguna.

Kondisi tersebut menunjukkan bahwa jumlah sentimen saja belum cukup untuk menggambarkan kualitas layanan secara menyeluruh. Oleh karena itu, penelitian ini melanjutkan analisis terhadap data negatif menggunakan pemodelan topik LDA untuk mengetahui aspek layanan yang menjadi sumber keluhan masyarakat.

6. Identifikasi Isu Kritis Menggunakan Latent Dirichlet Allocation (LDA)

Data dengan kategori sentimen negatif digunakan sebagai masukan dalam proses pemodelan topik menggunakan Latent Dirichlet Allocation (LDA). Tahapan ini bertujuan untuk menemukan pola isu yang sering muncul pada opini negatif masyarakat.

Penentuan jumlah topik dilakukan dengan membandingkan nilai coherence score dari beberapa variasi jumlah topik. Jumlah topik

dengan nilai coherence terbaik dipilih sebagai model akhir.

Jumlah Topik (K)	Coherence score
2	0,2277
3	0,2334
4	0,2751
5	0,2888
6	0,3256
7	0,3395
8	0,3315
9	0,3307
10	0,3265

Berdasarkan hasil pemodelan LDA, diperoleh beberapa kelompok isu kritis terkait layanan BPJS Kesehatan.

Contoh hasil interpretasi topik:

Topik	Kata Dominan	Interpretasi Isu
Topik 1	antrean, dokter, rumah sakit, pelayanan	Kendala pelayanan fasilitas kesehatan
Topik 2	aplikasi, mobile JKN, error, login	Permasalahan layanan digital
Topik 3	kartu, peserta, aktif, administrasi	Kendala administrasi kepesertaan
Topik 4	rujukan, faskes, prosedur	Hambatan proses rujukan

Hasil tersebut menunjukkan bahwa sentimen negatif masyarakat tidak hanya berasal dari satu aspek layanan, tetapi berkaitan dengan beberapa dimensi pelayanan BPJS Kesehatan.

7. Pembahasan

Hasil penelitian menunjukkan bahwa kombinasi IndoBERTweet dan LDA mampu memberikan informasi yang lebih komprehensif dibandingkan pendekatan analisis sentimen tunggal.

Model IndoBERTweet berhasil melakukan klasifikasi sentimen dengan accuracy 94%, menunjukkan bahwa model berbasis Transformer mampu memahami karakteristik bahasa informal pada media sosial X. Kemampuan tersebut dipengaruhi oleh proses continued fine-tuning yang memungkinkan model menyesuaikan representasi bahasa dengan domain layanan BPJS Kesehatan.

Berbeda dengan pendekatan berbasis machine learning tradisional yang bergantung pada fitur seperti TF-IDF, IndoBERTweet mampu memahami hubungan kontekstual antar kata

sehingga lebih sesuai untuk data media sosial yang memiliki variasi bahasa tinggi.

Selain mengetahui kecenderungan sentimen, penerapan LDA memberikan informasi tambahan mengenai faktor yang menyebabkan munculnya sentimen negatif. Hasil pemodelan menunjukkan bahwa isu layanan BPJS Kesehatan berkaitan dengan aspek pelayanan fasilitas kesehatan, layanan digital, administrasi kepesertaan, serta proses rujukan.

Temuan tersebut menunjukkan bahwa analisis berbasis NLP dapat digunakan sebagai pendekatan pendukung evaluasi layanan publik. Data opini masyarakat pada media sosial dapat dimanfaatkan sebagai sumber informasi tambahan untuk mengidentifikasi area layanan yang membutuhkan perhatian dan perbaikan.

5. KESIMPULAN

a. Penelitian ini berhasil menerapkan pendekatan Natural Language Processing (NLP) untuk menganalisis sentimen dan mengidentifikasi isu kritis layanan BPJS Kesehatan berdasarkan data cuitan pada media sosial X menggunakan IndoBERTweet dan Latent Dirichlet Allocation (LDA). Berdasarkan hasil penelitian terhadap 5.336 cuitan, model IndoBERTweet hasil continued fine-tuning menggunakan 1.000 data berlabel manual mampu melakukan klasifikasi sentimen dengan baik, ditunjukkan melalui nilai accuracy sebesar 94% pada data pengujian dengan distribusi precision, recall, dan F1-score yang seimbang pada masing-masing kelas sentimen. Hasil inferensi menunjukkan bahwa opini masyarakat terhadap layanan BPJS Kesehatan memiliki distribusi sentimen yang relatif seimbang, yaitu 2.673 cuitan positif (50,09%) dan 2.663 cuitan negatif (49,91%). Selanjutnya, penerapan Latent Dirichlet Allocation (LDA) terhadap data bersentimen negatif berhasil mengidentifikasi beberapa isu kritis yang berkaitan dengan layanan fasilitas kesehatan, layanan digital, administrasi kepesertaan, dan proses pelayanan. Hasil penelitian menunjukkan bahwa kombinasi IndoBERTweet dan LDA mampu memberikan informasi yang lebih komprehensif dibandingkan analisis sentimen tunggal karena tidak hanya mengetahui kecenderungan opini

masyarakat, tetapi juga mengungkap aspek layanan yang menjadi perhatian pengguna. Penelitian ini memiliki keterbatasan pada penggunaan data berbasis teks dari media sosial X dan belum mempertimbangkan data multimodal, sehingga penelitian selanjutnya dapat mengembangkan analisis dengan sumber data yang lebih beragam serta pendekatan model yang lebih kompleks.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi, serta seluruh pihak yang telah memberikan dukungan dan arahan dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] R. W. Harwenda, M. D. Angelo, I. Budi, A. B. Santoso, and P. K. Putra, "Sentiment Analysis on Government Public Policies: A Systematic Literature Review," *Dinasti Int. J. Educ. Manag. Soc. Sci.*, vol. 6, no. 5, pp. 4192–4211, 2025, doi: 10.38035/dijemss.v6i5.4699.
- [2] D. Darmawan *et al.*, "BPJS Patients Satisfaction Analysis Towards Service Quality of Public Health Center in Surabaya," *Media Kesehat. Masy. Indones.*, vol. 21, no. 2, pp. 107–112, 2022, [Online]. Available: <https://ejournal.undip.ac.id/index.php/mkmi>
- [3] M. Y. Al Qahar, Y. Ruldeviyani, U. N. Mukharomah, M. A. Fidyawan, and R. Putra, "Factor analysis influencing Mobile JKN user experience using sentiment analysis," *IAES Int. J. Artif. Intell.*, vol. 13, no. 2, pp. 1782–1793, 2024, doi: 10.11591/ijai.v13.i2.pp1782-1793.
- [4] D. Ramachandran and P. R., "Analysis of Twitter Specific Preprocessing Technique for Tweets," *Procedia Comput. Sci.*, vol. 165, pp. 245–251, 2020, doi: 10.1016/j.procs.2020.01.083.
- [5] N. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics*, vol. 9, no. 3, p. 483, 2020, doi: 10.3390/electronics9030483.
- [6] G. Alexander, M. Bahja, and G. F. Butt, "Automating Large-scale Health Care Service Feedback Analysis: Sentiment Analysis and Topic Modeling Study," *JMIR Med. Informatics*, vol. 10, no. 4, pp. 1–15, 2022, doi: 10.2196/29385.
- [7] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, "Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm," *J. Phys. Conf. Ser.*, vol. 1821, no. 1, 2021, doi: 10.1088/1742-6596/1821/1/012054.
- [8] S. Islam *et al.*, "A comprehensive survey on applications of transformers for deep learning tasks," *Expert Syst. Appl.*, vol. 241, 2024, doi: 10.1016/j.eswa.2023.122666.
- [9] F. Koto, J. Han lau, and T. Baldwin, "I NDO BERT WEET : A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," pp. 10660–10668, 2021.
- [10] L. Afuan, N. Hidayat, H. Hamdani, H. Ismanto, B. C. Purnama, and D. I. Ramdhani, "Optimizing BERT Models with Fine-Tuning for Indonesian Twitter Sentiment Analysis," *J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl.*, vol. 16, no. 2, pp. 248–267, 2025, doi: 10.58346/JOWUA.2025.12.016.
- [11] J. H. Jurafsky, D., & Martin, *Speech and Language Processing (3rd Edition Draft, 2021)*. Stanford University, 2021. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/>
- [12] H. M. Lee and Y. Sibaroni, "Comparison of IndoBERTweet and Support Vector Machine on Sentiment Analysis of Racing Circuit Construction in Indonesia," *J. Media Inform. Budidarma*, vol. 7, no. 1, p. 99, 2023, doi: 10.30865/mib.v7i1.5380.
- [13] E. Rijcken, K. Zervanov, P. Mosteiro, F. Scheepers, M. Spruit, and U. Kaymak, "Topic specificity: A descriptive metric for algorithm selection and finding the right number of topics," *Nat. Lang. Process. J.*, vol. 8, 2024, doi: 10.1016/j.nlp.2024.100082.
- [14] C. Weisser *et al.*, "Pseudo-document simulation for comparing LDA, GSDMM and GPM topic models on short and sparse text using Twitter data," *Comput. Stat.*, vol. 38, no. 2, pp. 647–674, 2023, doi: 10.1007/s00180-022-01246-z.
- [15] C. Schröer, F. Kruse, J. Marx, F. Kruse, and J. Marx, "ScienceDirect ScienceDirect A Systematic Literature Review A Systematic Literature Review on Applying Process Model on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.