

Deteksi Emosi Masyarakat Tentang Penyakit Gagal Ginjal Akut Dengan Metode EmoLex Dan Logistic Regression

^{1*}Wildan Gistra Irgamsyah, ²Risa Helilintar, ³Lilia Sinta Wahyuniar

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: *¹wildancancero@gmail.com, ²risa.helilintar@gmail.com, ³li2asint@gmail.com

Penulis Korespondens : Wildan Gistra Irgamsyah

Abstrak— Penelitian ini bertujuan untuk mendeteksi emosi masyarakat terhadap isu penyakit gagal ginjal akut dengan menggabungkan pendekatan leksikal berbasis EmoLex dan algoritma klasifikasi Logistic Regression. Data dikumpulkan dari media sosial Youtube tvOneNews. Setelah dilakukan tahap praproses data teks, setiap unggahan dipetakan ke dalam kategori emosi seperti marah, takut, sedih, dan lain-lain menggunakan leksikon NRC EmoLex. Selanjutnya, algoritma Logistic Regression digunakan untuk mengklasifikasikan emosi dominan berdasarkan fitur teks. Hasil penelitian menunjukkan bahwa emosi dominan yang muncul dalam diskusi publik adalah marah dan antisipasi, yang mencerminkan kekecewaan dan kewaspadaan masyarakat terhadap meningkatnya kasus gagal ginjal akut. Akurasi model mencapai 80%, dengan nilai precision dan recall tertinggi pada kategori emosi antisipasi dan marah. Penelitian ini memberikan gambaran awal tentang persepsi emosional masyarakat dalam isu krisis kesehatan, dan dapat dimanfaatkan oleh pemerintah atau lembaga kesehatan sebagai dasar pengambilan keputusan dalam strategi komunikasi publik.

Kata Kunci— Analisis Sentimen, Deteksi Emosi, EmoLex, Logistic Regression, Penyakit gagal ginjal akut

Abstract— This study aims to detect public emotions regarding the issue of acute kidney failure by combining a lexical approach based on EmoLex and the Logistic Regression classification algorithm. Data were collected from the social media platform YouTube, specifically from tvOneNews. After conducting text data preprocessing, each comment was mapped into emotion categories such as anger, fear, sadness, and others using the NRC EmoLex lexicon. Next, the Logistic Regression algorithm was used to classify the dominant emotion based on text features. The results show that the dominant emotions appearing in public discussions are anger and anticipation, reflecting public disappointment and vigilance regarding the rise in acute kidney failure cases. The model achieved an accuracy of 80%, with the highest precision and recall scores in the emotion categories of anticipation and anger. This study provides an initial overview of public emotional perception in a health crisis issue and can be utilized by the government or health institutions as a basis for decision-making in public communication strategies.

Keywords— Sentiment Analysis, Emotion Detection, EmoLex, Logistic Regression, Acute Kidney Injury

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Peningkatan kasus gagal ginjal akut di Indonesia, terutama pada anak-anak, menjadikan kondisi ini sebagai isu kesehatan yang serius dan berisiko tinggi terhadap keselamatan jiwa. Isu

ini menarik perhatian publik secara luas dan menjadi perbincangan di media sosial, mencerminkan kekhawatiran masyarakat terhadap aspek kesehatan dan keamanan obat. Media sosial memiliki peranan yang signifikan dalam kehidupan sosial masyarakat modern, khususnya dalam mendukung komunikasi lintas dunia, mempercepat penyebaran informasi, serta memengaruhi pembentukan opini publik. Media sosial berperan penting sebagai sumber ekspresi emosi masyarakat terhadap isu sosial dan kesehatan. Oleh karena itu, analisis emosi dari data teks di media sosial menjadi metode yang efektif untuk memahami persepsi dan reaksi masyarakat terhadap penyakit gagal ginjal akut.

Penelitian yang menggunakan leksikon emosi EmoLex yang diperluas dengan sinonim berhasil mendeteksi emosi pada teks media sosial, khususnya status Facebook, dengan akurasi sekitar 61,53%, berdasarkan klasifikasi delapan kategori emosi Plutchik dan sentimen, serta memungkinkan perluasan leksikon untuk meningkatkan cakupan kata[1]. Studi lain yang menerapkan EmoLex pada komentar akademik mahasiswa menunjukkan rata-rata akurasi klasifikasi emosi sebesar 56,2%, dengan tingkat akurasi tertinggi pada emosi sedih (72%), bahagia (68%), dan takut (68%), namun hasil tersebut dinilai masih belum memuaskan sehingga diperlukan pengembangan lebih lanjut terhadap leksikon ini[2]. Penelitian lain menunjukkan bahwa kombinasi algoritma Naïve Bayes dan Particle Swarm Optimization (PSO) efektif untuk klasifikasi emosi pada lirik lagu berbahasa Indonesia, dengan akurasi mencapai 90–96% [3]. Penelitian terhadap 96.485 komentar Instagram tentang Timnas Indonesia di Piala Asia 2023 menunjukkan dominasi sentimen positif (74,9%) dengan emosi utama *trust* dan *joy*, meskipun muncul kritik dan emosi negatif seperti *anger* dan *fear*[4]. Regresi logistik kerap digunakan sebagai metode klasifikasi emosi untuk memprediksi kemungkinan kemunculan emosi tertentu berdasarkan fitur yang diperoleh dari leksikon seperti EmoLex, dan meskipun kinerjanya cukup baik dalam klasifikasi biner maupun multikelas, hasilnya masih berada di bawah metode deep learning[5]. Secara keseluruhan, studi-studi menunjukkan bahwa pendekatan berbasis leksikon seperti EmoLex cukup efektif dalam menganalisis emosi pada teks berbahasa Indonesia, namun tingkat akurasi yang masih berada di kisaran 55–60% mendorong perlunya pengembangan metode klasifikasi lain, seperti regresi logistik atau ensemble learning, untuk meningkatkan kinerja[6].

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan emosi masyarakat Indonesia terkait isu gagal ginjal akut melalui data media sosial, serta mengevaluasi efektivitas metode kombinasi EmoLex dan Logistic Regression dalam pengklasifikasian emosi. Tujuan umum penelitian ini adalah untuk memberikan gambaran yang lebih detail mengenai persepsi emosi masyarakat terhadap isu tersebut, sementara tujuan spesifiknya adalah menguji akurasi metode yang digunakan dalam klasifikasi emosi. Hipotesis yang diajukan adalah bahwa kombinasi metode EmoLex dan Logistic Regression dapat meningkatkan akurasi deteksi emosi masyarakat dibandingkan metode analisis sentimen konvensional. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem analisis emosi berbasis teks berbahasa Indonesia serta memberikan dasar ilmiah untuk penyusunan strategi komunikasi kesehatan yang lebih efektif.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis teks berbasis data media sosial. Tujuan utamanya adalah untuk mendeteksi dan mengklasifikasikan emosi masyarakat Indonesia terhadap isu gagal ginjal akut. Subjek penelitian berupa komentar masyarakat di media sosial Youtube tvOneNews pada tanggal 20 Oktober 2022 yang berjudul “Gagal Ginjal Akut, Gegara Obat Sirup? Begini Kata Ketum PP IDAI”. Data dikumpulkan melalui proses scraping menggunakan youtube-comment-downloader, menghasilkan total 1.500 komentar yang dijadikan bahan penelitian.

Data yang terkumpul kemudian melalui proses pra-pemrosesan teks menggunakan teknik standar dalam Natural Language Processing (NLP). Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan yang fokus pada studi mengenai bagaimana manusia dan komputer berinteraksi melalui penggunaan bahasa alami[7]. Tahap prapemrosesan dilakukan untuk menyaring dan membersihkan data mentah yang telah dikumpulkan, sehingga data tersebut layak digunakan dalam tahap analisis atau pemrosesan selanjutnya. Tahapan prapemrosesan meliputi cleaning (pembersihan simbol, angka, dan tautan URL). *Case folding* merupakan proses mengubah seluruh huruf dalam teks menjadi satu format huruf, baik semua huruf kecil maupun besar, sesuai dengan kebutuhan analisis. Tokenisasi merupakan proses memecah teks atau kalimat menjadi unit-unit kecil yang disebut token untuk mempermudah analisis data teks. *Stopword removal* merupakan proses menghapus kata-kata umum yang tidak memiliki kontribusi penting terhadap makna dalam analisis teks suatu dataset. Normalisasi merupakan proses penyesuaian data ke dalam bentuk yang seragam atau standar untuk mempermudah tahap analisis selanjutnya[8]. Terakhir *Stemming* adalah proses untuk mengidentifikasi atau mengembalikan sebuah kata ke bentuk dasarnya[9]. *Stemming* menggunakan pustaka NLP Bahasa Indonesia seperti Sastrawi.

Setelah tahap pembersihan teks, setiap tweet dipetakan ke dalam kategori emosi menggunakan kamus leksikal NRC Emotion Lexicon (EmoLex). Kamus ini mengasosiasikan kata-kata dengan delapan kategori emosi dasar, yaitu: marah, takut, sedih, jijik, senang, percaya, antisipasi, dan terkejut, sebagaimana dijelaskan oleh Mohammad dan Turney (2013)[10]. Pemetaan dilakukan dengan menghitung frekuensi kemunculan kata-kata yang termasuk dalam masing-masing kategori emosi, dan emosi dominan dalam satu komentar ditentukan berdasarkan frekuensi tertinggi.

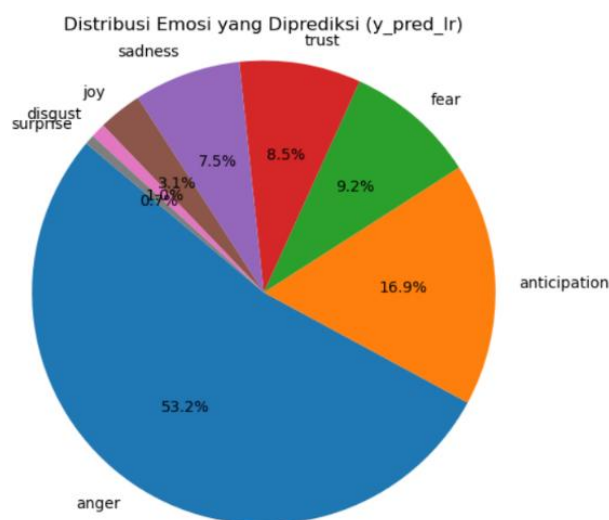
Selanjutnya, hasil pemetaan emosi digunakan sebagai data input untuk proses klasifikasi menggunakan model Logistic Regression, sebuah model statistik yang sering digunakan dalam klasifikasi biner dan multikelas. Dataset dibagi menjadi dua bagian, yaitu 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*). Model Logistic Regression dilatih untuk mengenali pola distribusi emosi berdasarkan fitur kata yang dimiliki masing-masing komentar.

Evaluasi performa model dilakukan menggunakan empat metrik evaluasi utama dalam klasifikasi teks, yaitu: akurasi, presisi (*precision*), daya tarik (*recall*), dan F1-score. Metrik-metrik ini memberikan gambaran menyeluruh tentang sejauh mana model mampu mengklasifikasikan emosi dengan benar dan konsisten. Semua proses analisis dilakukan

menggunakan bahasa pemrograman Python dengan pustaka seperti *scikit-learn*, *pandas*, dan *nlTK* untuk NLP.

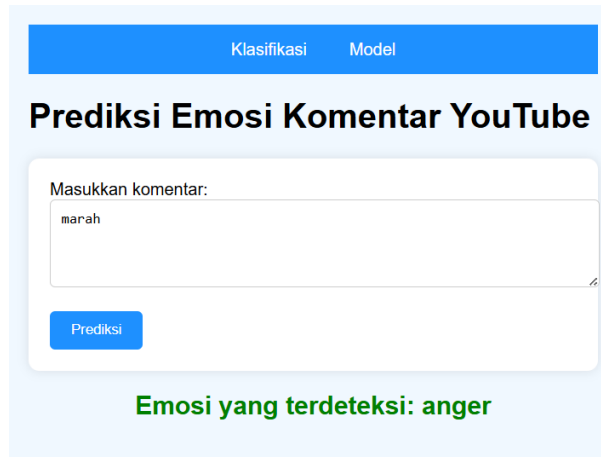
III. HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengumpulkan sebanyak 1.500 komentar dari media sosial Youtube Youtube tvOneNews pada tanggal 20 Oktober 2022 yang berjudul “Gagal Ginjal Akut, Gegara Obat Sirup? Begini Kata Ketum PP IDAI”. Analisis awal terhadap data menunjukkan bahwa puncak diskusi publik terjadi pada pertengahan Oktober, saat media mulai gencar memberitakan kasus gagal ginjal akut pada anak-anak. Fenomena ini mengindikasikan adanya hubungan kuat antara ekspos media dan respons emosional masyarakat yang tercermin dalam media sosial.



Gambar 1. Pie Chart Sebaran Emosi

Setelah dilakukan tahap pra-pemrosesan teks, setiap komentar dipetakan ke dalam delapan kategori emosi dasar menggunakan NRC Emotion Lexicon (EmoLex). Dalam kajian psikologi, emosi manusia umumnya diklasifikasikan ke dalam delapan kategori utama, yaitu marah (*anger*), takut (*fear*), antisipasi (*anticipation*), jijik (*disgust*), sedih (*sadness*), terkejut (*surprise*), yakin (*trust*), dan bahagia (*joy*) [11]. Hasil pemetaan emosi menunjukkan bahwa emosi yang paling dominan adalah rasa marah, yang muncul pada sekitar 53,2% dari total komentar. Emosi ini diikuti oleh rasa antisipasi (16,9%), takut (9,2%), dan percaya (8,5%), sementara sisanya mencerminkan emosi lain seperti sedih, senang, jijik, dan terkejut. Distribusi ini menunjukkan bahwa emosi negatif mendominasi persepsi publik terhadap isu ini. Dominasi rasa marah dan antisipasi dapat dijelaskan secara teoritis dengan merujuk pada teori emosi dasar Plutchik, yang menyebut bahwa emosi-emosi ini kerap muncul dalam situasi penuh ketidakpastian, ancaman terhadap kehidupan, dan rasa kehilangan. Dalam konteks ini, ketakutan terhadap keamanan obat serta kesedihan karena korban yang terlibat adalah anak-anak, menjadi pemicu utama respons emosional masyarakat.



Gambar 2. Antarmuka Aplikasi Deteksi Emosi

Hasil pemetaan emosi tersebut kemudian digunakan sebagai data untuk pelatihan model klasifikasi berbasis algoritma Logistic Regression. Logistic Regression merupakan metode analisis yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel prediktor[12]. Model ini dipilih karena kemampuannya yang cukup baik dalam menangani data dengan banyak fitur dan memberikan hasil klasifikasi yang dapat diinterpretasi dengan baik. Evaluasi performa model dilakukan dengan membagi data menjadi 80% data latih dan 20% data uji. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 80%, dengan nilai precision sebesar 0,77, recall sebesar 0,63, dan F1-score sebesar 0,68. Nilai-nilai ini mencerminkan keseimbangan antara ketepatan model dalam mendeteksi emosi yang benar dan kemampuannya dalam menangkap emosi yang relevan dari seluruh data yang tersedia. Namun demikian, terdapat indikasi bahwa distribusi emosi yang tidak seimbang antar kelas—khususnya dominasi beberapa emosi seperti marah dan antisipasi dapat memengaruhi kinerja model, khususnya dalam mengklasifikasikan emosi minoritas seperti jijik, senang, atau terkejut.

IV. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi, dapat disimpulkan bahwa pendekatan ini efektif dalam mengidentifikasi jenis-jenis emosi yang muncul di tengah masyarakat selama krisis kesehatan berlangsung. Temuan utama menunjukkan bahwa emosi dominan yang muncul adalah marah dan antisipasi, yang mengindikasikan tingginya tingkat kecemasan publik terhadap keselamatan anak-anak dan kepercayaan terhadap sistem kesehatan. Metode klasifikasi yang digunakan mampu mencapai tingkat akurasi yang cukup baik, sehingga dapat menjadi alternatif metodologi dalam konteks deteksi emosi berbasis data teks bahasa Indonesia.

Sebagai saran, penelitian ini menyarankan pengembangan lebih lanjut terhadap sumber daya leksikal yang digunakan, terutama kamus emosi berbahasa Indonesia, agar dapat merepresentasikan nuansa emosi lokal secara lebih akurat. Selain itu, disarankan untuk mengeksplorasi pendekatan algoritma lain yang mungkin memberikan hasil klasifikasi yang lebih optimal, khususnya dalam menangani data yang tidak seimbang. Penambahan variasi data dari platform media sosial lain juga direkomendasikan agar analisis emosi mencerminkan

persepsi masyarakat secara lebih luas dan beragam. Penelitian ini memberikan kontribusi terhadap pengembangan sistem analisis emosi dalam ranah teknik informasi dan data sains, serta membuka peluang kolaborasi interdisipliner dalam memahami dinamika sosial berbasis teks digital. Dengan demikian, pendekatan ini berpotensi menjadi alat bantu dalam mendukung pengambilan keputusan dan strategi komunikasi publik berbasis data yang lebih empatik dan responsif.

DAFTAR PUSTAKA

- [1] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *J. Eksplora Inform.*, vol. 9, no. 1, pp. 70–76, Sep. 2019, doi: 10.30864/EKSPLORA.V9I1.277.
- [2] A. Hamzah and R. Y. Ariyana, "Klasifikasi Emosi Berbasis Emolex dari Komentar Evaluasi Akademik Mahasiswa," *Techno.Com*, vol. 23, no. 2, pp. 455–466, May 2024, doi: 10.62411/TC.V23I2.10058.
- [3] M. Syauqi, A. Wijaya, W. Samudra, M. F. Z. F, A. G. Wp, and A. D. Kalifia, "KLASIFIKASI EMOSI PADA LIRIK LAGU MENGGUNAKAN ALGORITMA NAÏVE BAYES," *Sci. J. Ilm. Sains dan Teknol.*, vol. 3, no. 3, pp. 387-392–387–392, Jan. 2025, Accessed: Jun. 03, 2025. [Online]. Available: <https://jurnal.researchideas.org/index.php/scientica/article/view/182>
- [4] I. Made *et al.*, "Analisis Sentimen Berbasis Aspek dan Deteksi Emosi dari Ulasan Timnas Indonesia di Instagram," *JOINTECOMS (Journal Inf. Technol. Comput. Sci.)*, vol. 4, no. 4, pp. 273–284, Dec. 2024, doi: 10.47111/JOINTECOMS.V4I4.10792.
- [5] R. Nisa Sofia Amriza, D. Supriyadi, P. DI Jl Panjaitan No, K. Purwokerto Selatan, K. Banyumas, and J. Tengah, "Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media," *J. JUPITER*, vol. 13, no. 2, pp. 130–139.
- [6] S. F. Pane, F. Abdullah, and R. Habibi, "DETEKSI EMOSI PADA TEKS BERBAHASA INDONESIA MENGGUNAKAN PENDEKATAN ENSEMBLE," *JTT (Jurnal Teknol. Ter.)*, vol. 10, no. 2, pp. 80–90, Oct. 2024, doi: 10.31884/JTT.V10I2.551.
- [7] M. Furqan, S. Sriani, and M. N. Shidqi, "Chatbot Telegram Menggunakan Natural Language Processing," *Walisongo J. Inf. Technol.*, vol. 5, no. 1, pp. 15–26, Jun. 2023, doi: 10.21580/WJIT.2023.5.1.14793.
- [8] F. N. Zaman, M. A. Fadhilah, M. A. Ulinuha, and K. Umam, "MENGANALISIS RESPONS NETIZEN TWITTER TERHADAP PROGRAM MAKAN SIANG GRATIS MENERAPKAN NLP METODE NAIVE BAYES," *Just IT J. Sist. Informasi, Teknol. Informasi, dan Komput.*, vol. 14, no. 3, pp. 201–208, May 2024, doi: 10.24853/JUSTIT.14.3.201-208.
- [9] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *J. Transform.*, vol. 20, no. 2, pp. 1–12, Jan. 2023, doi: 10.26623/TRANSFORMATIKA.V20I2.5374.
- [10] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013, doi: 10.1111/J.1467-8640.2012.00460.X;WGROU:STRING:PUBLICATION.

- [11] R. Artikel, D. Caesy, R. #1, S. Khomsah, Y. Fathoni, and □ #3, “Emotions Analysis of Tourist Using Lexicon Text Analysis,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 13-27–13 – 27, May 2024, doi: 10.28932/JUTISI.V10I1.6690.
- [12] D. Nasien *et al.*, “Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes,” *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, Feb. 2024, doi: 10.58794/JEKIN.V4I1.640.