TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

Klasifikasi Sentimen Masyarakat Terhadap Kaesang Pangarep pada Media Sosial Twitter/X Menggunakan MLP Classifier dengan Fitur FastText

Veci Cahyono Tarmizi, Surya Agustian*, Okfalisa, Pizaini

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1]11950113466@students.uin-suska.ac.id, [2,*]surya.agustian@uin-suska.ac.id, [3]okfalisa@gmail.com, [4]pizaini@uin-suska.ac.id
Email Penulis Korespondensi: surya. agustian@uin-suska. ac. id

Abstrak–Media sosial kini menjadi wadah utama bagi masyarakat dalam mengekspresikan pandangan terhadap dinamika politik di Indonesia. Salah satu isu yang banyak diperbincangkan adalah pengangkatan Kaesang Pangarep sebagai Ketua Umum Partai Solidaritas Indonesia (PSI). Penelitian ini berupaya untuk menganalisis dan mengklasifikasikan sentimen publik terhadap isu tersebut dengan menerapkan algoritma Multi-Layer Perceptron (MLP) serta representasi teks menggunakan FastText. Data dikumpulkan melalui platform Twitter dengan memanfaatkan kata kunci seperti "Kaesang PSI", kemudian diperluas dengan tambahan data dari berbagai topik umum seperti Covid-19 dan Open Topic, agar distribusi sentimen positif, netral, dan negatif lebih proporsional serta mencerminkan opini publik yang lebih luas. Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, recall, dan F1 Score. Hasil pengujian menunjukkan bahwa model MLP-FastText memperoleh nilai berturut-turut F1 Score sebesar 0,5129; akurasi 0,6035; presisi 0,5319; dan recall 0,5996. Temuan ini mengindikasikan bahwa pendekatan gabungan antara MLP dan FastText memiliki kemampuan yang cukup baik dalam mengenali pola sentimen pada teks, khususnya pada konteks media sosial yang bersifat dinamis dan tidak terstruktur, serta menunjukkan efektivitas ketika dipadukan dengan strategi penambahan data eksternal yang relevan.

Kata Kunci: Analisis Sentimen; FastText; Kaesang Pangarep; Multi-Layer Perceptron; Twitter

Abstract–Social media has become a primary channel for the public to express their opinions and reactions toward various political developments in Indonesia. One of the prominent discussions revolves around Kaesang Pangarep's appointment as the Chairman of the Indonesian Solidarity Party (PSI). This study aims to analyze and classify public sentiment regarding this issue by employing the Multi-Layer Perceptron (MLP) algorithm integrated with FastText-based text representation. The dataset was collected from Twitter using keywords such as "Kaesang PSI", and was further expanded with additional data from general topics including Covid-19 and Open Topic, ensuring a balanced distribution across positive, neutral, and negative sentiment categories for a more comprehensive representation of public opinion. The model's performance was evaluated through four metrics: accuracy, precision, recall, and F1 Score. The experimental results demonstrate that the MLP–FastText model achieved consecutive scores of 0. 5129 for F1 Score, 0. 6035 for accuracy, 0. 5319 for precision, and 0. 5996 for recall. These findings indicate that the combination of MLP and FastText effectively captures sentiment patterns within textual data, particularly in the context of unstructured and dynamic social media content, and performs well when enhanced with relevant external data augmentation strategies.

Keywords: Sentiment Analysis; FastText; Kaesang Pangarep; Multi-Layer Perceptron; Twitter

1. PENDAHULUAN

Kaesang Pangarep resmi menjadi anggota Partai Solidaritas Indonesia (PSI) pada Sabtu, 23 September 2023, dengan menerima kartu tanda anggota yang diserahkan langsung oleh pengurus PSI di Kota Solo, Jawa Tengah. PSI, yang didirikan pada akhir 2014, merupakan salah satu partai politik baru di Indonesia dengan visi pembangunan nasional yang berlandaskan pada nilai demokrasi, kemanusiaan, keberagaman, keadilan, kemajuan, dan martabat bangsa. Kaesang, putra bungsu Presiden Joko Widodo, kemudian ditetapkan sebagai Ketua Umum PSI, keputusan yang menimbulkan berbagai tanggapan publik karena keluarga Presiden selama ini dikenal dekat dengan Partai Demokrasi Indonesia Perjuangan (PDIP). Kehadirannya di PSI dinilai mampu menarik perhatian pemilih muda, khususnya generasi Z yang cenderung kurang aktif dalam politik, sekaligus membawa perspektif baru yang banyak diperbincangkan di media sosial seperti Twitter (Rohman, 2023)

Twitter telah menjadi salah satu sumber data utama dalam penelitian analisis sentimen dan opini publik karena jutaan pengguna global berbagi pemikiran mereka setiap hari, yang menyediakan teks real-time yang beragam untuk dianalisis dengan teknik NLP dan pembelajaran mesin (Qi & Shabrina, 2023); (Wang et al., 2022)

Tingginya popularitas Twitter menjadikan platform ini sebagai sarana bagi masyarakat untuk mengekspresikan pendapat dan pandangan publik terhadap isu yang tengah hangat dibahas (Wijaya et al., 2021). Analisis sentimen merupakan teknik dalam bidang *Natural Language Processing*(NLP) yang digunakan untuk memahami sekaligus mengidentifikasi perasaan, emosi, serta opini yang terdapat dalam suatu teks. Teknik ini berfungsi untuk menentukan kecenderungan sentimen terhadap suatu fenomena, apakah bersifat positif, negatif, atau netral (Yusanto & Akbar, 2024). Analisis sentimen adalah teknik analisis berbasis teks yang digunakan di media sosial dan platform internet lainnya untuk mengumpulkan data. Dengan demikian, analisis sentimen dapat menjadi sumber data yang berguna untuk memahami polaritas opini seseorang terhadap sebuah objek, apakah opini tersebut positif atau negatif. Dalam penelitian ini, metode analisis sentimen menggunakan Natural Language Processing (NLP) untuk memproses dan menganalisis teks (Wisnu et al., 2020).

Banyak metode yang digunakan untuk klasifikasi sentiment salah satunya adalah *Multi Layer Perceptron (MLP) Classifier* yan merupakan salah satu algoritma Artificial Neural Network (ANN) (Baihaqi et al., 2023). Keuntungan

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

utama dari MLP adalah potensi pembelajaran yang tinggi, ketahanan terhadap noise, nonlinier, paralelisme, toleransi kesalahan, dan kemampuan yang tinggi dalam menggeneralisasi tugas (Heidari et al., 2020). MLP memiliki keunggulan dalam mempelajari pola nonlinier, tahan terhadap noise, serta mampu melakukan generalisasi dengan baik. Namun, dalam konteks pemrosesan bahasa alami, MLP umumnya bekerja dengan representasi fitur berbasis vektor seperti TF-IDF atau word embedding statis, sehingga tidak secara eksplisit mempertimbangkan urutan kata dalam teks. Berbeda dengan model sekuensial seperti LSTM atau Transformer yang dirancang untuk menangkap hubungan antar kata berdasarkan urutan kemunculannya, MLP lebih menekankan pada pola global dalam data (Yang et al., 2023).

Pada penelitian tentang menggabungkan sentistrength dan multilayer perceptron dalam klasifikasi sentimen twitter menghasilkan penambahan data pelatihan dapat meningkatkan akurasi rata-rata sebesar 5% dari akurasi awal. Multilayer Perceptron lebih akurat daripada Naive Bayes dengan rasio akurasi tertinggi 77,71% dan 76,07% (Prastowo et al., 2019). Dalam penelitian tentang analisis algoritma multilayer perceptron pada sentimen pengguna twitter terhadap vaksin covid-19 menyebutkan bahwa hasil performa dari penelitian ini mendapatkan akurasi tertinggi sebesar 81. 2%, presisi sebesar 83. 8%, dan recall sebesar 71. 2%. Hasil analisis sentimen dapat dilihat pada opini masyarakat terhadap vaksin COVID-19 yang terbagi menjadi tiga kategori: 35% opini positif, 16,3% opini negatif, dan 48,7% opini netral (Pasaribu et al., 2023). Penelitian pada dataset kecil mengunaakan MLP Classifier dan TF-IDF dapat mengatasi keterbatasan dataset dan mencapai F1-score sebesar 0,6767 dan akurasi 0,6667(Arasy et al., 2025).

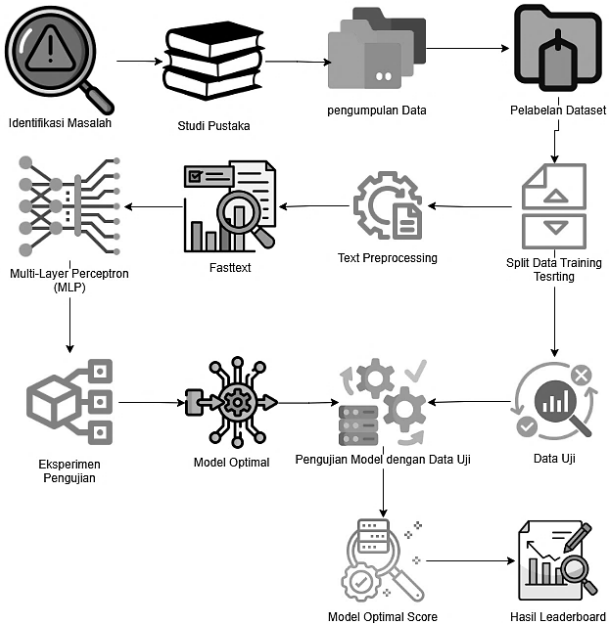
Dalam penelitian ini, selain memanfaatkan model *Multi-Layer Perceptron (MLP)*, digunakan juga fitur *FastText* sebagai model word embedding yang telah dilatih sebelumnya menggunakan jumlah teks yang besar dari berbagai sumber. *FastText* mampu menangkap penggunaan kata dalam berbagai bahasa, idiom, serta konteks yang berbeda, sehingga model dapat mengenali hubungan semantik dan asosiasi antar kata berdasarkan kemunculannya dalam teks (Ardinata et al., 2024). Penelitian terdahulu melakukan klasifikasi sentimen terkait vaksin Covid-19 di Twitter dengan memanfaatkan algoritma *K-Nearest Neighbor* dan *FastText*. Data diperoleh melalui Python dan Twitter API, kemudian dilabeli menggunakan metode crowdsourcing dan majority voting. Setelah proses penyeimbangan, dataset terdiri dari 6. 000 data untuk training, 778 untuk development, dan 400 untuk testing, dengan hasil terbaik mencapai akurasi 69% dan F1-score 60% (Naldi & Agustian, 2023). Penelitian lain mengenai Peningkatan Klasifikasi Otomatis Kalimat dalam Latihan Ilmu Data menunjukkan bahwa pengklasifikasi MLP dengan embedding FastText memberikan hasil terbaik. Saat menggunakan embedding BERT, akurasi dan F1-score MLP sedikit menurun, namun tetap lebih unggul dibanding pengklasifikasi lain (Angelone et al., 2022).

Berdasarkan uraian latar belakang tersebut, peneliti tertarik untuk menganalisis sentimen pengguna Twitter terhadap Kaesang Pangarep sebagai Ketua Umum Partai Solidaritas Indonesia (PSI). Oleh karena itu, penelitian ini mengambil judul "Klasifikasi Sentimen Masyarakat Di *Twitter* Terhadap Kaesang Pangarep Sebagai Ketua Umum Partai Solidaritas Indonesia Dengan *Multi-Layer Perceptron (MLP)* Fitur *Fasttext"*.

2. METODOLOGI PENELITIAN

2. 1 Tahapan Penelitian

Tahapan penelitian adalah rangkaian langkah-langkah yang digunakan sebagai acuan dalam melaksanakan sebuah penelitian. Adapun tahapan-tahapan penelitian yang dilakukan ditampilkan pada Gambar 1 sebagai berikut:



Gambar 1. Tahapan Penelitian

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

2. 2 Identifikasi Masalah

Perumusan masalah merupakan tahap awal dalam metodologi penelitian, di mana peneliti bertujuan untuk mengidentifikasi, merumuskan, dan memahami permasalahan yang akan diteliti. Pada tahap ini, peneliti mengamati fenomena yang terjadi, mengumpulkan informasi awal, serta mengkaji persoalan secara mendalam agar dapat menentukan fokus penelitian yang jelas dan spesifik.

Selanjutnya, peneliti menelaah latar belakang masalah untuk menemukan hubungan antara fenomena yang diamati dengan teori atau konsep yang relevan. Proses ini memungkinkan peneliti menyusun pertanyaan penelitian yang tepat, sehingga hasil penelitian nantinya dapat memberikan kontribusi ilmiah yang signifikan dan solusi yang aplikatif terhadap permasalahan yang diangkat.

2. 3 Studi Pustaka

Pada tahap ini, peneliti melakukan pengumpulan dan kajian terhadap berbagai literatur serta konsep dasar yang relevan untuk membangun kerangka teori penelitian. Sumber informasi diperoleh dari beragam media, antara lain buku referensi, jurnal ilmiah, prosiding konferensi, tesis atau disertasi, serta konten digital seperti video edukatif di YouTube dan sumber daring lainnya. Pemilihan sumber dilakukan secara selektif untuk memastikan kualitas dan relevansi materi yang mendukung fokus penelitian.

Setelah dikumpulkan, seluruh informasi dianalisis dan disintesiskan secara kritis untuk mengidentifikasi kesenjangan penelitian, memperkuat argumen, dan mendukung metodologi yang diterapkan. Proses ini bertujuan agar setiap konsep dan teori yang digunakan memiliki dasar yang jelas serta dapat menjadi landasan kuat dalam pelaksanaan penelitian, sehingga hasil yang diperoleh nantinya memiliki validitas dan reliabilitas yang tinggi.

2. 4 Pengumpulan Data

Pengumpulan data (data crawling) dalam penelitian ini dilakukan secara otomatis menggunakan bahasa pemrograman Python. Proses pengambilan dan analisis data memanfaatkan beberapa pustaka Python, yaitu Tweepy untuk pengambilan data dari Twitter serta Scikit-learn untuk implementasi algoritma Multi Layer Perceptron (MLP) dalam klasifikasi sentimen. Pengambilan data dilakukan menggunakan sejumlah kata kunci yang berkaitan dengan topik penelitian, antara lain “Kaesang Pangarep”, “Kaesang Pangarep Ketua Umum PSI”, “Ketua Umum PSI”, dan “Partai Solidaritas Indonesia”. Selain itu, peneliti menambahkan kata kunci lain yang relevan dengan isu terpilihnya Kaesang Pangarep sebagai Ketua Umum PSI guna memperoleh data yang lebih luas dan beragam. Pendekatan ini bertujuan agar data yang dikumpulkan mencerminkan persepsi publik terhadap isu politik yang berkembang di media sosial, khususnya Twitter

Selain mengumpulkan data utama yang berhubungan dengan isu politik tersebut, penelitian ini juga menambahkan sekitar 10. 000 hingga 15. 000 tweet dari berbagai topik umum yang tidak memiliki keterkaitan langsung dengan isu utama. Penambahan data eksternal ini dilakukan agar model yang dibangun tidak bersifat bias terhadap satu topik tertentu dan memiliki kemampuan generalisasi yang lebih baik. Proses pengumpulan data dilaksanakan dalam rentang waktu 25 September 2023 hingga 3 Oktober 2023. Dengan jangka waktu tersebut, data yang diperoleh diharapkan dapat memberikan gambaran yang komprehensif mengenai opini publik terhadap fenomena sosial-politik yang terjadi pada periode tersebut. .

2. 5 Pelabelan Dataset

Pelabelan dataset dilakukan dengan menggunakan metode crowdsourcing labelling yang membagi data ke dalam tiga kategori utama, yaitu positif, netral, dan negatif. Metode ini dipilih karena mampu memproses data dalam jumlah besar secara lebih efisien dengan melibatkan partisipasi beberapa individu dalam memberikan label. Setiap individu diminta untuk membaca teks hasil crawling kemudian menentukan kategori sentimennya berdasarkan konteks kalimat. Proses pelabelan ini menjadi salah satu tahapan penting karena hasil label akan memengaruhi performa model klasifikasi yang dikembangkan.

Meskipun metode crowdsourcing efektif dalam menangani volume data besar, kualitas hasil pelabelan dapat bervariasi tergantung pada tingkat pemahaman dan konsistensi penilai. Untuk meminimalkan potensi kesalahan, penelitian ini menggunakan pendekatan majority voting, di mana setiap data diberi label oleh beberapa orang, kemudian label akhir ditentukan berdasarkan suara terbanyak (Agustian et al., 2024). Pendekatan ini meningkatkan reliabilitas hasil pelabelan dan mengurangi pengaruh subjektivitas individu. Dengan cara ini, dataset yang dihasilkan menjadi lebih konsisten dan layak digunakan dalam proses pelatihan model klasifikasi sentimen.

2. 6 Text Preprocessing

Text preprocessing adalah langkah awal dalam pengolahan data teks sebelum dilakukan analisis lebih lanjut atau pembangunan model. Tahap ini bertujuan untuk membersihkan teks mentah dari elemen-elemen yang tidak diperlukan dan mengubahnya menjadi bentuk yang lebih terstruktur serta konsisten. Dengan penerapan text preprocessing, data menjadi lebih seragam dan lebih mudah diolah oleh algoritma pembelajaran mesin.

Tahapan ini sangat penting untuk meningkatkan akurasi model, karena data yang belum bersih dapat menimbulkan kesalahan interpretasi dan hasil klasifikasi yang kurang optimal. Dalam penelitian ini, proses text preprocessing dilakukan melalui beberapa langkah utama yang dirancang agar data siap digunakan sebagai input pada

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

model klasifikasi sentimen secara efektif. Adapun tahapan text preprocessing yang diterapkan dalam penelitian ini meliputi beberapa langkah utama yaitu:

a. Case folding
   Case folding adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase) untuk memastikan konsistensi dalam analisis data teks. Langkah ini dilakukan agar perbedaan antara huruf kapital dan huruf kecil tidak memengaruhi hasil pemrosesan, sehingga kata seperti "Data" dan "data" diperlakukan sebagai entitas yang sama, mengurangi redundansi dan potensi kesalahan interpretasi. Selain itu, case folding memudahkan algoritma pembelajaran mesin dalam mengenali pola kata dengan lebih akurat, terutama ketika data berasal dari berbagai sumber yang memiliki format penulisan tidak konsisten. Tahapan ini merupakan bagian penting dari text preprocessing karena meningkatkan kualitas input bagi model analisis atau klasifikasi, sehingga berkontribusi pada peningkatan akurasi dan efisiensi proses pengolahan data (Andika et al., 2019).
b. Cleaning
   Cleaning merupakan tahap pembersihan teks yang bertujuan untuk menghapus karakter khusus, tanda baca, angka, atau simbol-simbol lain yang tidak relevan dan berpotensi mengganggu proses analisis data. Proses ini penting dilakukan agar data teks menjadi lebih bersih dan terstruktur, sehingga algoritma pembelajaran mesin dapat memproses informasi secara lebih akurat. Dengan melakukan tahap cleaning, redundansi dan gangguan dari elemen-elemen tidak penting dapat diminimalkan, sehingga kualitas data input meningkat dan hasil analisis atau klasifikasi menjadi lebih handal (Susilawati & Iqbal, 2025).
c. Emoji Handling
   Emoji handling adalah proses pengelolaan emoji dalam data teks sebelum dilakukan pemrosesan lebih lanjut, seperti pelatihan model machine learning atau analisis sentimen. Tahap ini dilakukan untuk memastikan bahwa simbol-simbol emoji, yang sering menyampaikan emosi atau makna tambahan dalam komunikasi digital, dapat diinterpretasikan dengan tepat oleh algoritma. Dengan menangani emoji secara sistematis, data teks menjadi lebih konsisten dan informasi emosional yang terkandung dalam emoji dapat dimanfaatkan untuk meningkatkan akurasi analisis sentimen atau performa model prediktif (Khan et al., 2025).

2. 7 Fasttext

FastText merupakan sebuah library sumber terbuka yang dikembangkan oleh Facebook AI Research (FAIR) untuk mendukung berbagai kebutuhan dalam bidang pemrosesan bahasa alami (Natural Language Processing). Library ini digunakan untuk menghasilkan word embedding, yaitu representasi kata dalam bentuk vektor numerik yang memungkinkan sistem komputer memahami makna kata berdasarkan konteks kemunculannya. Pendekatan ini membantu mesin mengenali hubungan semantik antar kata, sehingga dapat digunakan dalam berbagai aplikasi seperti klasifikasi teks, analisis sentimen, dan sistem rekomendasi berbasis teks (Kuyumcu et al., 2019; Umer et al., 2023). Dengan kemampuannya dalam memproses bahasa alami secara efisien, FastText telah menjadi salah satu alat yang banyak digunakan dalam penelitian dan pengembangan model berbasis teks.

Keunggulan utama dari FastText terletak pada kemampuannya dalam memanfaatkan karakter n-gram sebagai dasar pembentukan vektor kata. Dengan menyimpan informasi dalam potongan karakter, metode ini memungkinkan model menghasilkan representasi vektor bahkan untuk kata yang tidak ditemukan secara langsung dalam kamus pelatihan atau out-of-vocabulary words. Fitur tersebut menjadikan FastText lebih unggul dibandingkan metode embedding konvensional seperti Word2Vec, karena mampu menangani variasi morfologis, bentuk turunan kata, maupun kesalahan ejaan tanpa kehilangan makna semantik dari kata tersebut (Kuyumcu et al., 2019; Umer et al., 2023).

2. 8 *Multi-Layer Perceptron* (MLP)

Multi-Layer Perceptron (MLP) adalah jaringan yang terdiri dari perceptron. Jaringan ini memiliki lapisan input yang menerima sinyal input, lapisan output yang membuat prediksi atau keputusan untuk input yang diberikan, dan lapisan yang ada di antara lapisan input dan lapisan output disebut lapisan tersembunyi. Ada banyak lapisan tersembunyi tersembunyi, jumlah lapisan tersembunyi dapat diubah sesuai kebutuhan (Goh & Yann, 2021). Multilayer Perceptron (MLP) adalah kelas feedforward jaringan. Dengan kata lain, disebut sebagai jaringan syaraf tiruan nonlinier dan kuat. Pada tahun 2015 diusulkan MLP terdiri dari tiga lapisan, yaitu lapisan input, lapisan tersembunyi dan akhirnya lapisan keluaran. Setiap simpul atau lapisan berisi neuron-neuron dengan fungsi aktivasi nonlinier. Dan sebagai tambahan, MLP menggunakan estimator aktivasi universal yang berisi satu simpul tersembunyi dan memiliki beberapa unit nonlinier yang berbeda. Dan untuk alasan itu, hubungan pembelajaran dari variabel yang baik satu sama lain. Dalam Multilayer Perceptron (MLP) aliran data searah dari input ke lapisan keluaran. Lapisannya yang bermacam-macam dan fungsi aktivasi nonlinier untuk membedakan MLP dari perceptron linier (Ali Kandhro et al., 2019).

2. 9 Pengujian

Kinerja model dievaluasi dengan memanfaatkan *confusion matrix*, yang digunakan untuk menghitung tingkat akurasi serta mengukur sejauh mana model mampu melakukan prediksi dengan benar (Alnuaim et al., 2022). Model diimplementasikan berdasarkan rancangan eksperimen yang telah ditetapkan sebelumnya. Dataset kemudian dibagi ke dalam tiga bagian utama, yaitu data pelatihan untuk membangun model, data validasi untuk menyesuaikan parameter, serta data pengujian untuk mengevaluasi performa akhir. Melalui eksperimen ini, diharapkan dapat diperoleh gambaran

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

yang lebih menyeluruh mengenai kinerja Multi-Layer Perceptron (MLP) yang dipadukan dengan fitur FastText dalam melakukan analisis sentimen, sekaligus memberikan pemahaman yang lebih dalam mengenai efektivitas model dalam berbagai konteks penerapan.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Dataset

Dalam penelitian ini, dataset yang digunakan mencakup tiga kelompok topik utama, yaitu Kaesang, Covid-19, dan Open Topic. Ketiga topik tersebut berfungsi sebagai dasar penting dalam meningkatkan kinerja serta akurasi model pada tahap pelatihan dan evaluasi. Pada topik Kaesang, data dibagi menjadi dua bagian, yakni Kaesang0 dengan jumlah 240 data dan Kaesang1 dengan 300 data. Dari keseluruhan data Kaesang1, sebanyak 20% atau sekitar 60 data dialokasikan sebagai data validasi. Dengan demikian, total data latih gabungan untuk topik Kaesang berjumlah 480 data, sedangkan data validasi mencapai 120 data.

Sementara itu, pada topik Covid-19 dilakukan proses undersampling untuk menyeimbangkan jumlah data dengan topik lainnya, hingga diperoleh 510 data yang siap digunakan. Proses undersampling ini bertujuan agar model tidak mengalami bias terhadap kelas dengan jumlah data yang lebih besar. Untuk topik Open Topic, dilakukan proses sampling hingga terkumpul sebanyak 380 data yang representatif. Secara keseluruhan, pengumpulan dan pengolahan dataset ini dilakukan dengan hati-hati agar distribusi data di setiap topik tetap seimbang dan mampu mendukung proses pelatihan model secara optimal. Data yang digunakan dalam penelitian ini ditampilkan pada Tabel 1, Tabel 2, dan Tabel 3 sebagai sampel dari masing-masing topik.

Tabel 1. Tabel Kaesang

No	Kalimat	Label
1	Kaesang di PSI? Harapan baru bagi kaum muda dan partai! Pasti banyak ide segar.	Positif
2	Kaesang jadi Ketum PSI. Sekarang fokusnya konsolidasi dan Pemilu 2024. Kita lihat saja nanti.	Netral
3	Kaesang jadi Ketum PSI ini mendadak banget. Jangan-jangan cuma manuver politik sesaat.	Negatif

Tabel 2. Data Covid

No	Kalimat	Label
1	Vaksinasi berhasil! Kasus COVID-19 turun, aktivitas kembali normal. Lega rasanya.	Positif
2	Pemerintah terus pantau COVID-19 dan varian baru. Protokol kesehatan tetap diupdate	Netral
3	COVID-19 masih ada, tapi banyak yang cuek. Khawatir gelombang baru datang lagi.	Negatif

Tabel 2. Data Open Topic

No	Kalimat	Label
1	Belajar online fleksibel banget, bisa dari mana saja. Mandiri dan melek digital.	Positif
2	Pembelajaran online butuh internet dan alat yang memadai. Akses merata masih jadi PR	Netral
3	Belajar online bikin cepat bosan, susah paham, dan kurang interaksi. Nggak efektif	Negatif

3.2 Teks Preprocessing

Proses pembersihan teks pada dataset tweet merupakan tahap krusial untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik dan relevan dengan tujuan penelitian. Tahapan ini dilakukan dengan menghapus berbagai elemen yang tidak berkontribusi terhadap hasil analisis, seperti URL, spasi berlebih, hashtag, serta username atau mention yang tidak memiliki makna informatif terhadap konteks sentimen. Dengan demikian, data mentah yang awalnya tidak terstruktur dapat diubah menjadi bentuk yang lebih siap untuk dianalisis secara komputasional.

Setelah pembersihan dasar dilakukan, tahap selanjutnya melibatkan penerapan beberapa teknik lanjutan untuk memperbaiki kualitas teks secara keseluruhan. Salah satu di antaranya adalah case folding, yaitu proses mengubah seluruh karakter huruf menjadi huruf kecil agar format teks menjadi seragam. Selain itu, karakter-karakter yang tidak diperlukan juga dihapus untuk mengurangi noise dalam data. Penanganan emoji juga menjadi bagian penting, karena simbol-simbol tersebut sering kali mengandung ekspresi emosional yang relevan dengan analisis sentimen.

Secara keseluruhan, kombinasi langkah-langkah tersebut menghasilkan teks yang lebih terstruktur, bersih, dan konsisten sehingga mampu meningkatkan efektivitas algoritma dalam proses pelatihan model machine learning. Hasil akhir dari tahap preprocessing ini disajikan pada Tabel 4, yang memperlihatkan bentuk data setelah melalui proses pembersihan menyeluruh dan siap digunakan pada tahap analisis berikutnya.

Tabel 3. *Text Preprocessing*

No	Proses	Sebelum	Sesudah
1	Case Folding	Kaesang di PSI? Harapan baru bagi kaum muda dan partai! Pasti banyak ide segar	kaesang di psi? harapan baru bagi kaum muda dan partai! pasti banyak ide segar

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

No	Proses	Sebelum	Sesudah
2	Text Cleaning	kaesang di psi? harapan baru bagi kaum muda dan partai! pasti banyak ide segar	kaesang di psi harapan baru bagi kaum muda dan partai pasti banyak ide segar
3	Emoji Handling	kaesang di psi harapan baru bagi kaum muda dan partai pasti banyak ide segar	kaesang di psi harapan baru bagi kaum muda dan partai pasti banyak ide segar

3.3 Eperiment Text Preprocessing

Serangkaian eksperimen text preprocessing dilakukan secara sistematis untuk menilai sejauh mana setiap teknik berkontribusi terhadap peningkatan performa model MLP Classifier. Dalam penelitian ini, berbagai metode preprocessing seperti case folding, text cleaning, dan emoji handling diuji secara terpisah maupun dalam kombinasi tertentu untuk memperoleh pemahaman yang lebih mendalam mengenai pengaruh masing-masing langkah terhadap hasil klasifikasi. Setiap kombinasi teknik diproses dan diuji menggunakan parameter yang sama, sehingga hasil yang diperoleh dapat dibandingkan secara objektif tanpa adanya bias dari faktor eksternal.

Tujuan utama dari eksperimen ini adalah untuk menentukan teknik preprocessing mana yang paling efektif dalam menghasilkan representasi teks yang bersih, informatif, dan konsisten, sehingga mampu meningkatkan kinerja model pada berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Melalui pendekatan ini, proses pengujian tidak hanya berfokus pada peningkatan performa model, tetapi juga pada pemahaman mendalam terhadap dampak setiap tahapan pembersihan data terhadap hasil klasifikasi.

Hasil lengkap dari eksperimen tersebut disajikan pada Tabel 5, yang menampilkan perbandingan performa antar kombinasi preprocessing secara komprehensif. Tabel ini memberikan gambaran jelas mengenai kontribusi tiap teknik terhadap peningkatan kualitas data teks dan efektivitas model MLP Classifier dalam melakukan analisis sentimen secara optimal.

Tabel 4. Eksperimen Preprocessing

Skema	Case Folding	Text Cleaning	Emoji Handling
S1	Ya	Ya	Ya
S2	Ya	Tidak	Tidak
S3	Ya	Tidak	Ya
S4	Tidak	Tidak	Ya
S5	Ya	Ya	Tidak
S6	Tidak	Ya	Ya
S7	Tidak	Ya	Tidak
S8	Tidak	Tidak	Tidak

Tabel 5 berisi hasil pengujian terhadap beberapa variasi langkah preprocessing yang digunakan dalam penelitian ini. Pada tabel tersebut, simbol “YES” mengindikasikan bahwa langkah tertentu diterapkan dalam proses preprocessing, sedangkan “NO” menunjukkan bahwa langkah tersebut tidak diikutsertakan.

3.4 Pengujian data

Tahap ini melibatkan pelaksanaan sejumlah eksperimen yang bertujuan untuk mengevaluasi performa model MLP Classifier melalui penerapan berbagai kombinasi parameter yang disusun secara terencana dan sistematis. Setiap skema eksperimen, yang diberi label E1 hingga E8, memiliki konfigurasi berbeda untuk menguji pengaruh variasi parameter terhadap hasil klasifikasi. Parameter yang dimodifikasi meliputi jumlah serta ukuran hidden layer, fungsi aktivasi yang digunakan, algoritma solver, nilai alpha sebagai pengatur regularisasi, learning rate, hingga jumlah iterasi maksimum (max iteration) yang menentukan batas proses pelatihan model.

Pendekatan ini bertujuan untuk menemukan konfigurasi paling optimal yang mampu memberikan performa terbaik pada model dalam hal akurasi dan stabilitas hasil. Dengan membandingkan setiap kombinasi parameter, dapat diketahui bagaimana setiap elemen mempengaruhi kemampuan model dalam mengenali pola pada data. Detail lengkap dari masing-masing skema eksperimen disajikan pada Tabel 6 Eksperimen Machine Learning, yang menampilkan konfigurasi parameter secara rinci sebagai dasar untuk analisis lebih lanjut terhadap kinerja model.

Tabel 5. Eksperimen Pengujian

Skema	Hidden Layer	Activation	Solver	Alpha	Learning Rate	Max Lter
E1	1024, 512, 256	relu	adam	0. 0001	0. 001	200
E2	800, 400, 200	tanh	sgd	0. 001	0. 01	300
E3	750, 500, 250	logistic	lbfgs	0. 01	0. 1	100
E4	1024, 256, 64	relu	adam	0. 0001	0. 001	400
E5	600, 400, 200	tanh	sgd	0. 001	0. 01	500
E6	512, 256, 128	logistic	lbfgs	0. 01	0. 1	200
E7	700, 350, 100	relu	adam	0. 0001	0. 001	300
E8	900, 450, 150	tanh	sgd	0. 001	0. 01	400
E9	10. 5	relu	adam	0. 0001	0. 001	300

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

Berdasarkan variasi parameter pada masing-masing skema, eksperimen ini bertujuan untuk menemukan konfigurasi terbaik yang menghasilkan performa model MLP Classifier paling optimal dalam melakukan klasifikasi data.

3.5 Hasil

Tabel berikut menyajikan hasil evaluasi dari kombinasi berbagai skema preprocessing (S1–S8) dan skema model MLP Classifier (E1–E9) yang diterapkan pada dataset tweet dengan dua kategori utama, yaitu Covid dan Open Topic. Setiap kombinasi diuji secara sistematis untuk mengetahui bagaimana variasi pada tahap preprocessing serta konfigurasi model memengaruhi kinerja klasifikasi. Evaluasi dilakukan menggunakan dua metrik utama, yakni F1 Score dan Accuracy, yang dianggap mampu memberikan gambaran menyeluruh terhadap keseimbangan antara presisi dan recall dari model yang digunakan. Selain itu, jumlah data yang digunakan dalam setiap eksperimen juga diperhitungkan agar perbandingan antar kombinasi tetap konsisten dan valid.

Meskipun dalam tabel tidak ditampilkan secara eksplisit, dataset Kaesang juga terlibat dalam keseluruhan eksperimen ini. Dataset tersebut digunakan secara konsisten pada setiap skema, baik dari sisi preprocessing maupun konfigurasi model, sehingga kontribusinya tetap ada dalam proses pelatihan dan evaluasi. Namun, karena jumlah data Kaesang yang digunakan selalu sama di setiap skema, dataset ini tidak dimuat secara terpisah dalam tabel agar penyajian hasil tetap ringkas dan fokus pada perbandingan antar kombinasi preprocessing serta model.

Secara keseluruhan, hasil yang diperoleh dari berbagai kombinasi menunjukkan bahwa variasi preprocessing dan parameter model memiliki pengaruh signifikan terhadap performa klasifikasi sentimen. Beberapa kombinasi tertentu mampu menghasilkan nilai F1 Score dan Accuracy yang lebih tinggi dibandingkan yang lain, menandakan bahwa pemilihan teknik preprocessing yang tepat serta konfigurasi parameter model yang optimal sangat penting untuk meningkatkan efektivitas MLP Classifier dalam mengolah teks dan mengenali pola sentimen dengan lebih akurat. Berikut ditampilkan pada Tabel 7 hasil evaluasi pada penelitian ini.

Tabel 6. Hasil Evaluasi

No	Skema Pre Processing	Skema Machine Learning	Dataset						Total Data	F1 Score	Accuracy	Rank F1 Score	Rank Accuracy
			Covid			Open Topic							
			Positif	Netral	Negatif	Positif	Netral	Negatif					
1	S1	E1	35	45	40	50	40	45	255	0.54	0. 58	38	39
…	…	…	…	…	…	…	…	…	…	…	…	…	…
9	S1	E9	10	50	60	20	70	20	230	0.65	0. 67	6	2
10	S2	E1	350	300	400	260	250	270	1830	0.54	0. 59	38	31
…	…	…	…	…	…	…	…	…	…	…	…	…	…
61	S7	E7	45	40	55	35	35	35	245	0.52	0. 58	46	39
62	S7	E8	130	115	155	100	100	90	690	0.69	0. 67	1	2
63	S7	E9	9	34	55	80	81	56	315	0.64	0. 63	8	10
…	…	…	…	…	…	…	…	…	…	…	…	…	…
72	S8	E9	29	80	80	12	52	33	286	0.55	0. 57	34	46
73	S8	E6	0	0	0	0	0	0	0	0.00	0. 32	73	73

Berdasarkan hasil pengujian yang ditampilkan pada Gambar 4, diperoleh bahwa model MLP Classifier yang digunakan berhasil mencapai nilai F1 Score sebesar 0. 5129, akurasi sebesar 0. 6035, presisi sebesar 0. 5319, dan recall sebesar 0. 5996. Hasil ini menunjukkan bahwa model mampu mencapai keseimbangan yang cukup baik antara kemampuan mengenali sentimen positif maupun negatif (recall) serta ketepatan dalam melakukan klasifikasi (precision). Kombinasi nilai-nilai metrik tersebut menggambarkan bahwa model dapat bekerja secara stabil dalam mengidentifikasi pola sentimen pada data teks yang dianalisis.

Dengan capaian akurasi yang melampaui 60%, model dapat dikatakan memiliki performa yang cukup andal dalam mengklasifikasikan sentimen pada topik yang berkaitan dengan pengangkatan Kaesang sebagai Ketua Umum PSI. Meskipun masih terdapat ruang untuk peningkatan performa, hasil ini sudah menunjukkan bahwa pendekatan dan teknik preprocessing yang diterapkan mampu memberikan kontribusi positif terhadap kinerja model. Secara keseluruhan, hasil pengujian ini mengindikasikan bahwa model MLP Classifier memiliki potensi yang baik yang ditampilkan pada Tabel 8 untuk digunakan dalam analisis sentimen media sosial dengan topik serupa.

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

Tabel 7. Hasil Leaderboard

Tim	Metode	Rank	F1 Score	Acc	Prec	Recall
Rank 1 (Pranata et al., 2025)	BERT	1	0. 6063	0. 7053	0. 5936	0. 6775
Organizer (El Saputra et al., 2024)	SVM	14	0. 5128	0. 6121	0. 5289	0. 5722
Abdurrahman Arrasy (Arasy et al., 2025)	MLP + TF-IDFVectorizer	51	0. 5027	0. 5915	0. 5327	0. 5851
Admin (Agustian et al., 2024)	BASELINE	21	0. 4038	0. 4545	0. 4953	0. 488
Penelitian ini	MLP + FASTEXT	13	0. 5129	0. 6035	0. 5319	0. 5996

Berdasarkan hasil yang ditampilkan pada Tabel 8, metode BERT menempati posisi peringkat pertama (Rank 1) dengan performa terbaik di antara seluruh model yang diuji. Model ini berhasil mencapai F1 Score sebesar 0. 6063 dan akurasi sebesar 0. 7053, dengan keseimbangan yang baik antara precision (0. 5936) dan recall (0. 6775). Nilai tersebut menunjukkan bahwa BERT mampu mengenali dan mengklasifikasikan kelas sentimen dengan tingkat akurasi yang tinggi dan konsistensi yang baik. Kemampuan representasi kontekstual dari BERT memungkinkan model ini memahami makna kalimat secara lebih mendalam, sehingga menghasilkan performa yang unggul dibandingkan metode lainnya.

Di sisi lain, metode Support Vector Machine (SVM) yang digunakan oleh tim Organizer menempati peringkat ke-14, dengan F1 Score sebesar 0. 5128 dan akurasi sebesar 0. 6121. Meskipun performanya sedikit lebih rendah dibandingkan BERT, SVM tetap menunjukkan hasil yang cukup layak dalam konteks klasifikasi sentimen. Perbedaan hasil ini mengindikasikan bahwa meskipun SVM merupakan metode klasik yang masih banyak digunakan karena efisiensi dan kestabilannya, model berbasis deep learning seperti BERT mampu memberikan peningkatan performa yang signifikan dalam memahami konteks bahasa alami.

Sementara itu, model baseline yang dikembangkan oleh tim Admin menempati peringkat ke-21, dengan performa paling rendah di antara model yang diuji, yaitu F1 Score sebesar 0. 4038 dan akurasi sebesar 0. 4545, serta precision 0. 4953 dan recall 0. 488. Hal ini menunjukkan bahwa model baseline belum mampu menangani kompleksitas klasifikasi sentimen dengan baik. Adapun model yang digunakan dalam penelitian ini, yaitu MLP + FastText, berhasil mencapai peringkat ke-13, sedikit lebih baik dibandingkan SVM milik Organizer. Model ini memperoleh F1 Score sebesar 0. 5129, akurasi 0. 6035, precision 0. 5319, dan recall 0. 5996, yang menandakan bahwa pendekatan yang digunakan cukup kompetitif dan mampu bersaing dengan metode konvensional, bahkan mendekati performa model-model yang lebih kompleks seperti BERT. .

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode Multi-Layer Perceptron (MLP) dengan pendekatan FastText untuk melakukan klasifikasi sentimen terhadap topik pengangkatan Kaesang Pangarep sebagai Ketua Umum Partai Solidaritas Indonesia (PSI). Berdasarkan hasil evaluasi, model MLP dengan FastText menunjukkan performa yang cukup kompetitif dengan nilai F1 Score, akurasi, presisi, dan recall berturut-turut sebesar 0,51; 0,60; 0,53; dan 0,60. Capaian ini menempatkan model pada peringkat ke-13 dari seluruh peserta kompetisi dan sedikit lebih unggul dibandingkan model SVM milik penyelenggara (organizer) yang menempati peringkat ke-14. Hasil ini menunjukkan bahwa kombinasi MLP dan FastText mampu memberikan kinerja yang solid dalam klasifikasi sentimen, terutama pada konteks data media sosial yang cenderung bervariasi dan tidak terstruktur. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Salah satu keterbatasan utama terletak pada jumlah dan distribusi data yang tidak merata, terutama pada skema eksperimen yang melibatkan jumlah data terbatas atau ketidakseimbangan antar kelas sentimen. Kondisi tersebut dapat memengaruhi stabilitas hasil dan kemampuan model dalam melakukan generalisasi terhadap data baru. Selain itu, performa terendah yang diperoleh pada skema S8-E6, dengan F1 Score 0,00 dan akurasi hanya 0,32, menunjukkan bahwa penggunaan data yang tidak lengkap berdampak signifikan terhadap efektivitas model. Oleh karena itu, diperlukan perbaikan pada tahap pengumpulan dan penyeimbangan data untuk meningkatkan keandalan model di masa mendatang. Untuk penelitian mendatang, direkomendasikan penggunaan dataset dengan ukuran yang lebih luas, mencerminkan representasi data yang baik, serta memiliki keseimbangan antar kelas, sehingga model dapat mempelajari pola dengan lebih efektif dan memberikan hasil yang lebih stabil. Selain itu, penerapan metode klasifikasi berbasis deep learning seperti BERT layak dipertimbangkan, mengingat metode tersebut terbukti unggul dengan capaian akurasi tertinggi sebesar 0. 70 dan F1 Score 0. 60 dalam eksperimen yang dilakukan. Penelitian ini sendiri memberikan kontribusi awal yang menjanjikan dalam pengembangan sistem klasifikasi sentimen berbasis MLP dan FastText, serta membuka peluang bagi penelitian selanjutnya untuk mengeksplorasi teknik yang lebih canggih dan adaptif guna meningkatkan akurasi serta kemampuan generalisasi model dalam memahami konteks bahasa alami secara lebih mendalam.

REFERENCES

Agustian, S., et al. (2024). *New directions in text classification research: Maximizing the performance of sentiment classification from limited data*. https://github.com/s4gustian/Small_DataSet_Sentiment_Classification

Agustian, S., Syah, M. I., Fatiara, N., & Abdillah, R. (2024). *New directions in text classification research: Maximizing the performance of sentiment classification from limited data*. arXiv. https://arxiv.org/abs/2407.05627

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

Ali Kandhro, I., Chhajro, M. A., Kumar, K., Lashari, H. N., & Khan, U. (2019). Student feedback sentiment analysis model using various machine learning schemes: A review. *Indian Journal of Science and Technology, 14*(12), 1–9. https://doi.org/10.17485/ijst/2019/v12i14/143243

Alnuaim, A. A., et al. (2022). Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier. *Journal of Healthcare Engineering, 2022*, 1–12. https://doi.org/10.1155/2022/6005446

Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis sentimen masyarakat terhadap hasil quick count pemilihan presiden Indonesia 2019 pada media sosial Twitter menggunakan metode naive Bayes classifier. *Indonesian Journal of Applied Statistics, 2*(1), 34. https://doi.org/10.13057/ijas.v2i1.29998

Angelone, A. M., Galassi, A., & Vittorini, P. (2022). Improved automated classification of sentences in data science exercises. In F. De la Prieta et al. (Eds.), *Methodologies and intelligent systems for technology enhanced learning* (pp. 12–21). Springer. https://doi.org/10.1007/978-3-030-86618-1_2

Arasy, A., Agustian, S., Handayani, L., & Iskandar, I. (2025). Klasifikasi sentimen menggunakan metode multilayer perceptron dengan fitur TF-IDF. *MALCOM: Indonesian Journal of Machine Learning and Computer Science, 5*(3), 908–919. https://doi.org/10.57152/malcom.v5i3.2052

Ardinata, P. M. S., Permana, A. A. J., Wijaya, I. N. S. W., Teknik, F., & Kejuruan, D. (2024). Identifikasi dan normalisasi teks slang dengan FastText pada Twitter dalam bahasa Indonesia. *Jurnal Pendidikan Teknologi dan Kejuruan, 21*(1). https://doi.org/10.23887/jptkundiksha.v21i1.66381

Baihaqi, W. M., Yunita, I. R., Damayanti, A. S. T., & Akhaerunnisa, L. (2023). Analysis of classification algorithm performance on user review sentiment of the Muamalat DIN application. *CogITo Smart Journal, 9*(2), 241–251. https://doi.org/10.31154/cogito.v9i2.511.241-251

El Saputra, Y., Agustian, S., Yusra, Y., & Ramadhani, S. (2024). Klasifikasi sentimen SVM dengan dataset yang kecil pada kasus Kaesang sebagai ketua umum PSI. *KLIK: Kajian Ilmiah Informatika dan Komputer, 4*(6), 2902–2908. https://doi.org/10.30865/klik.v4i6.1944

Goh, A. M., & Yann, X. L. (2021). A novel sentiments analysis model using perceptron classifier. *International Journal of Electronics Engineering and Applications, 10*(2), 1–10. https://doi.org/10.30696/IJEEA.IX.IV.2021.01-10

Heidari, A. A., Faris, H., Mirjalili, S., Aljarah, I., & Mafarja, M. (2020). Ant lion optimizer: Theory, literature review, and application in multi-layer perceptron neural networks. In S. Mirjalili, J. Song Dong, & A. Lewis (Eds.), *Nature-inspired optimizers: Theories, literature reviews and applications* (pp. 23–46). Springer. https://doi.org/10.1007/978-3-030-12127-3_3

Khan, A., Majumdar, D., & Mondal, B. (2025). Sentiment analysis of emoji fused reviews using machine learning and BERT. *Scientific Reports, 15*(1), Article 7538. https://doi.org/10.1038/s41598-025-92286-0

Kuyumcu, B., Aksakalli, C., & Delil, S. (2019). An automated new approach in fast text classification (fastText). In *Proceedings of the 3rd International Conference on Natural Language Processing and Information Retrieval* (pp. 1–4). ACM. https://doi.org/10.1145/3342827.3342828

Naldi, A., & Agustian, S. (2023). Klasifikasi sentimen vaksin COVID-19 menggunakan K-nearest neighbor berdasarkan word embeddings FastText pada Twitter. *ZONAsi: Jurnal Sistem Informasi, 5*(2), 323–333. https://doi.org/10.31849/zn.v5i2.12548

Pasaribu, F. H., Khairina, N., Noviandri, D., Susilawati, S., & Syah, R. (2023). Analysis of the multilayer perceptron algorithm on Twitter user's sentiment towards the COVID-19 vaccine. *Journal of Informatics and Telecommunication Engineering, 7*(1), 155–163. https://doi.org/10.31289/jite.v7i1.9664

Pranata, J., Agustian, S., Jasril, J., & Haerani, E. (2025). Penggunaan model bahasa indoBERT pada metode random forest untuk klasifikasi sentimen dengan dataset terbatas. *Building of Informatics, Technology and Science (BITS), 6*(3), 1668–1676. https://doi.org/10.47065/bits.v6i3.6335

Prastowo, E. Y., Endroyono, & Yuniarno, E. M. (2019). Combining SentiStrength and multilayer perceptron in Twitter sentiment classification. In *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)* (pp. 381–386). IEEE. https://doi.org/10.1109/ISITIA.2019.8937134

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining, 13*(1), Article 31. https://doi.org/10.1007/s13278-023-01030-x

Rohman, N. (2023). Peran Partai Solidaritas Indonesia (PSI) dalam pemilihan presiden 2024: Analisis terhadap pemilih pemula. *JPW (Jurnal Politik Walisongo), 5*(1), 85–102. https://doi.org/10.21580/jpw.v5i1.18330

Susilawati, S., & Iqbal, M. (2025). Penerapan metode Naïve Bayes untuk mengidentifikasi sentimen pengguna pada ulasan aplikasi ReelShort di Google Play Store. *SIMKOM, 10*(1), 49–59. https://doi.org/10.51717/simkom.v10i1.686

Umer, M., et al. (2023). Impact of convolutional neural network and FastText embedding on text classification. *Multimedia Tools and Applications, 82*(4), 5569–5585. https://doi.org/10.1007/s11042-022-13459-x

Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment analysis of Twitter data. *Applied Sciences, 12*(22), Article 11775. https://doi.org/10.3390/app122211775

Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis sentimen opini publik tentang Undang-Undang Cipta Kerja pada Twitter. *Jambura Journal of Electrical and Electronics Engineering, 3*(2), 78–83. https://doi.org/10.37905/jjeee.v3i2.10885

TIN: Terapan Informatika Nusantara
Vol 6, No 7, December 2025, page 1105-1114
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i7.8815

Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes. In *Journal of Physics: Conference Series* (Vol. 1444, No. 1, Article 012034). Institute of Physics Publishing. https://doi.org/10.1088/1742-6596/1444/1/012034

Yang, E., et al. (2023). Transformer versus traditional natural language processing: How much data is enough for automated radiology report classification? *British Journal of Radiology*, *96*(1149). https://doi.org/10.1259/bjr.20220769

Yusanto, Y., & Akbar, M. (2024). Analisis sentimen Jogja darurat sampah di Twitter menggunakan ekstraksi fitur model Word2Vec dan convolutional neural network. *TIN: Terapan Informatika Nusantara*, *4*(10), 679–688. https://doi.org/10.47065/tin.v4i10.4952