

Analisis Sentimen Opini Publik terhadap BPI Danantara di Media Sosial X Menggunakan IndoRoBERTa dan SVM

Firnanda Putra Agustinus¹, Moch Afaldo Danny Ardiansyah²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri
Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹[*1firnandanocoment224@gmail.com](mailto:firnandanocoment224@gmail.com), ²aldodanny3@gmail.com

Abstrak – Media sosial menjadi sumber utama opini masyarakat terhadap isu publik seperti Danantara, namun data yang dihasilkan bersifat besar, tidak terstruktur, dan mengandung variasi bahasa yang kompleks. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap isu Danantara dengan membandingkan metode *Support Vector Machine* (SVM) dan model IndoRoBERT. Penelitian menggunakan pendekatan kuantitatif dengan tahapan pengumpulan data media sosial, prapemrosesan teks meliputi pembersihan data, tokenisasi, dan normalisasi, kemudian ekstraksi fitur menggunakan TF-IDF untuk pemodelan SVM. Selanjutnya, pemodelan berbasis IndoRoBERT dilakukan untuk memahami konteks semantik teks secara lebih mendalam. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Hasil menunjukkan distribusi sentimen relatif konsisten antara kedua model, dengan accuracy SVM terhadap pseudo-label IndoRoBERTa mencapai 98% precision dan recall rata-rata 0.98, macro F1-score 0.97. IndoRoBERTa terbukti lebih stabil dalam menangani teks informal dan kontekstual. Penelitian ini menunjukkan bahwa model berbasis *transformer* lebih efektif digunakan untuk analisis sentimen berbahasa Indonesia pada isu publik strategis berbasis media sosial.

Kata Kunci — Analisis Sentimen, IndoRoBERTa, Media Sosial, SVM

1. PENDAHULUAN

Media sosial telah menjadi ruang utama bagi masyarakat dalam menyampaikan opini, kritik, dan penilaian terhadap berbagai isu publik. Besarnya jumlah pengguna serta kemudahan dalam berinteraksi menjadikan media sosial sebagai sumber data opini publik yang sangat kaya, namun bersifat tidak terstruktur dan terus berkembang. Komentar pengguna di media sosial, khususnya Instagram, merefleksikan tingkat kepercayaan publik terhadap suatu isu dan dapat dianalisis secara sistematis menggunakan pendekatan analisis sentimen berbasis *Natural Language Processing* untuk memperoleh gambaran persepsi masyarakat secara objektif dan terukur [1].

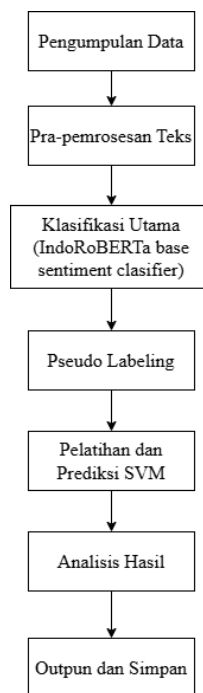
Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengklasifikasikan opini ke dalam kategori positif, negatif, atau netral berdasarkan data teks. Penerapan analisis sentimen pada data daring terbukti efektif dalam mengekstraksi persepsi pengguna terhadap suatu isu atau layanan, khususnya pada data berbahasa Indonesia yang bersumber dari platform digital. Analisis sentimen mampu menggambarkan persepsi pengguna secara sistematis meskipun data yang digunakan bersifat tidak terstruktur, terbatas, dan tidak seimbang [2]. Temuan serupa juga dikemukakan dalam penelitian [3] yang menunjukkan bahwa analisis sentimen media sosial mampu merepresentasikan opini publik secara objektif dan terukur melalui pengelompokan sentimen positif, negatif, dan netral pada data teks daring. Sedangkan, penelitian [4] mengungkapkan bahwa teks media sosial berbahasa Indonesia memiliki tingkat kompleksitas yang tinggi akibat penggunaan bahasa tidak baku, singkatan, serta makna kontekstual yang bergantung pada situasi, sehingga dibutuhkan pendekatan pemodelan yang mampu memahami konteks kalimat secara lebih mendalam dalam analisis sentimen.

Dalam lima tahun terakhir, pendekatan analisis sentimen di Indonesia berkembang dari penggunaan machine learning klasik menuju model deep learning berbasis transformer. Algoritma *Support Vector Machine* (SVM) dengan representasi fitur berbasis TF-IDF masih mampu memberikan kinerja klasifikasi sentimen yang cukup stabil pada data ulasan digital berbahasa Indonesia, meskipun keterbatasannya terlihat pada pemahaman konteks semantik yang kompleks seperti opini ambigu dan ironi [5]. Keterbatasan tersebut mendorong adopsi model berbasis *deep learning*, khususnya IndoBERT yaitu model bahasa pra-latih khusus bahasa Indonesia dan terbukti unggul dalam memahami konteks dua arah pada berbagai tugas NLP [6]. IndoBERT secara konsisten menghasilkan akurasi lebih tinggi dibandingkan metode klasik dalam analisis sentimen isu publik di media sosial karena kemampuannya menangkap makna kontekstual dan relasi semantik yang lebih dalam [7]. Oleh karena itu, kajian ini memposisikan SVM dan IndoBERT sebagai dua pendekatan representatif untuk dibandingkan dalam

menganalisis sentimen masyarakat terhadap isu publik strategis, sehingga dapat diperoleh gambaran objektif mengenai efektivitas masing-masing metode pada konteks data media sosial berbahasa Indonesia.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *text mining* untuk menganalisis sentimen masyarakat terhadap isu Danantara berdasarkan data media sosial. Klasifikasi sentimen dilakukan menggunakan *IndoBERT* sebagai pengklasifikasi awal untuk menghasilkan *pseudo label*, kemudian *Support Vector Machine* (SVM) digunakan pada tahap lanjutan karena terbukti efektif dalam klasifikasi sentimen teks berbahasa Indonesia[6][8]. Berikut adalah tahapan yang dilakukan dalam penelitian ini:



Gambar 1. Metologi Penelitian

2.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengambil data teks dari media sosial yang relevan dengan isu Danantara. Data dikumpulkan menggunakan teknik crawling berbasis API dengan kata kunci tertentu agar diperoleh opini publik yang representatif dan relevan terhadap topik penelitian. Pendekatan ini umum digunakan dalam penelitian analisis sentimen berbasis Twitter untuk menangkap dinamika opini masyarakat secara aktual [9]

2.2 Pra-pemrosesan Teks

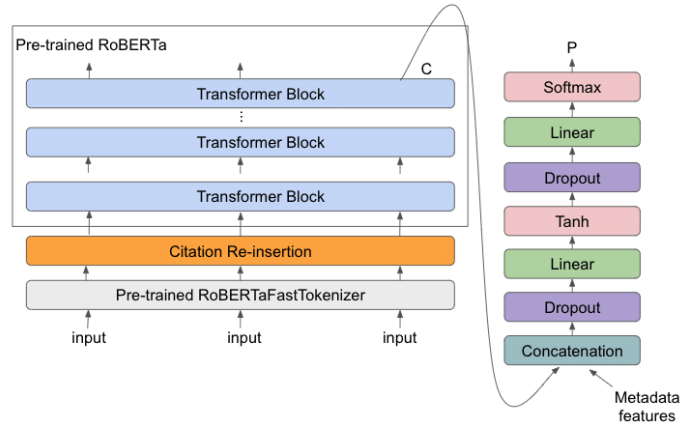
Pra-pemrosesan data merupakan bagian krusial dalam analisis sentimen yang bertujuan untuk menyiapkan data teks mentah sebelum dilakukan proses ekstraksi fitur menggunakan metode TF-IDF. Pada tahap ini, dilakukan pembersihan teks untuk menghilangkan karakter, simbol, dan elemen non-informatif yang tidak memiliki kontribusi terhadap makna sentimen. Teks dipecah menjadi unit kata melalui proses tokenisasi, diikuti dengan penghapusan stopword guna mengurangi dominasi kata-kata umum yang tidak merepresentasikan informasi sentimen secara signifikan. Proses pra-pemrosesan kemudian dilengkapi dengan stemming untuk mengubah kata ke bentuk dasarnya, sehingga variasi kata dapat diminimalkan dan representasi fitur menjadi lebih konsisten. Hasil dari seluruh tahapan ini digunakan sebagai masukan pada proses ekstraksi fitur berbasis TF-IDF untuk mendukung klasifikasi sentimen secara lebih akurat [10].

2.3 Klasifikasi Utama (*IndoRoBERTa Base Sentiment Classifier*)

Pada tahap ini, model IndoRoBERTa digunakan sebagai pengklasifikasi utama untuk memprediksi sentimen awal pada data teks. Model berbasis *transformer* ini dipilih karena kemampuannya dalam menangkap konteks bahasa Indonesia secara lebih baik dibandingkan metode berbasis fitur konvensional, sehingga menghasilkan prediksi sentimen yang lebih kontekstual, sebagaimana dibuktikan pada berbagai studi *transformer-based sentiment analysis* untuk teks berbahasa Indonesia [11].

2.4 Pseudo Labeling

Hasil prediksi sentimen dari model IndoRoBERTa selanjutnya digunakan sebagai *pseudo-label* pada data yang belum memiliki label manual. Pendekatan *pseudo-labeling* memungkinkan pemanfaatan data tidak berlabel secara lebih optimal untuk memperkaya data latih, khususnya ketika ketersediaan data berlabel terbatas, karena teknik ini memanfaatkan prediksi model awal untuk memberikan label semu pada sampel tidak berlabel dalam kerangka *semi-supervised learning* [12].



Gambar 2. Proses Pseudo-Labeling IndoRoBERTa untuk Pelatihan SVM

2.5 Pelatihan dan Prediksi SVM

Pada tahap ini, data teks yang telah melalui proses pra-pemrosesan dan memiliki label, baik label manual maupun pseudo-label, digunakan sebagai data latih dan data uji untuk membangun model klasifikasi sentimen. Sebelum proses pelatihan dilakukan, data teks ditransformasikan ke dalam bentuk representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* agar dapat diproses oleh algoritma klasifikasi berbasis *machine learning*, sebagaimana dijelaskan dalam penelitian yang menilai pengaruh teknik representasi teks terhadap akurasi klasifikasi sentimen teks.[13]

2.5.1 Ekstraksi Fitur TF-IDF

Metode *Term Frequency–Inverse Document Frequency (TF-IDF)* digunakan untuk memberikan bobot pada setiap term dalam dokumen dengan mempertimbangkan frekuensi kemunculan term tersebut pada suatu dokumen serta distribusinya pada seluruh dokumen dalam korpus. Pendekatan ini efektif dalam menurunkan pengaruh kata-kata umum yang kurang informatif dan meningkatkan daya pembeda fitur dalam klasifikasi sentimen berbasis teks tidak terstruktur [14].

Nilai *Term Frequency (TF)* menyatakan frekuensi kemunculan suatu term t pada dokumen d dan dirumuskan sebagai

$$TF(t, d) = \frac{f(t, d)}{\sum f(d)} \dots \dots \dots (1)$$

di mana $f(t, d)$ adalah jumlah kemunculan term t dalam dokumen d , dan $\sum f(d)$ merupakan total jumlah kata dalam dokumen tersebut. Selanjutnya, *Inverse Document Frequency (IDF)* digunakan untuk mengurangi bobot term yang muncul pada banyak dokumen dan dirumuskan sebagai:

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \dots \dots \dots (2)$$

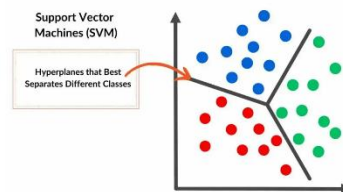
N sebagai jumlah total dokumen dalam korpus dan $df(t)$ sebagai jumlah dokumen yang mengandung term t . Bobot akhir TF-IDF diperoleh dengan mengalikan nilai TF dan IDF, sehingga dirumuskan sebagai:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \dots \dots \dots (3)$$

Hasil dari proses ekstraksi fitur TF-IDF berupa vektor numerik berdimensi tinggi yang merepresentasikan setiap dokumen dan selanjutnya digunakan sebagai masukan pada tahap pelatihan model klasifikasi sentimen menggunakan *Support Vector Machine (SVM)*.

2.5.2 Pelatihan Model Support Vector Machine

Support Vector Machine (SVM) digunakan sebagai algoritma klasifikasi untuk memisahkan data sentimen ke dalam kelas yang telah ditentukan. SVM bekerja dengan mencari sebuah *hyperplane* optimal yang mampu memaksimalkan margin antar kelas dalam ruang fitur berdimensi tinggi.



Gambar 3. Ilustrasi Pemisahan Kelas oleh Support Vector Machine

Fungsi keputusan SVM dapat dinyatakan sebagai:

$$f(x) = w \cdot x + b \dots \dots \dots (4)$$

x merupakan vektor fitur hasil TF-IDF, w adalah vektor bobot, dan b adalah bias. Proses pelatihan SVM bertujuan untuk menemukan parameter w dan b dengan meminimalkan fungsi objektif berikut:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \dots \dots \dots (5)$$

dengan batasan:

$$y_i(w \cdot x_i + b) \geq 1$$

x_i adalah vektor fitur dokumen ke- i dan y_i merupakan label kelas sentimen. Pendekatan ini bertujuan untuk menghasilkan margin maksimum sehingga model memiliki kemampuan generalisasi yang baik terhadap data baru [15].

2.5.3 Analisis Hasil

Hasil klasifikasi dianalisis untuk melihat distribusi sentimen serta pola opini masyarakat terhadap isu Danantara. Analisis ini bertujuan untuk memahami kecenderungan sentimen positif dan negatif, sekaligus mengevaluasi kemampuan model dalam menangkap karakteristik opini publik, sebagaimana diterapkan dalam studi-studi analisis sentimen opini masyarakat berbasis media sosial untuk mengukur persepsi publik terhadap isu tertentu

2.6 Output dan Simpan

Tahap akhir berupa penyajian hasil klasifikasi sentimen dalam bentuk visualisasi atau ringkasan statistik, kemudian disimpan untuk keperluan analisis lanjutan dan dokumentasi penelitian. Output ini menjadi dasar dalam penarikan kesimpulan dan pembahasan hasil penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari media sosial dan dikumpulkan berdasarkan kata kunci yang berkaitan dengan isu Danantara. Seluruh data yang terkumpul bersifat teks tidak terstruktur dan mengandung berbagai variasi bahasa informal khas media sosial. Oleh karena itu, sebelum dilakukan proses klasifikasi, data teks terlebih dahulu melalui tahap pra-pemrosesan yang mencakup pembersihan teks, normalisasi, penghapusan kata tidak relevan, serta penyesuaian format agar siap diproses oleh model klasifikasi. Tahapan ini bertujuan untuk mengurangi noise dan meningkatkan kualitas data sebelum masuk ke tahap pemodelan.

3.2 Pra-pemrosesan Teks

Pada gambar 4. menampilkan perbandingan data teks sebelum dan sesudah tahap pra-pemrosesan. Kolom *full_text* berisi teks mentah hasil pengumpulan data media sosial yang masih mengandung variasi penulisan, huruf kapital, dan elemen tidak relevan, sedangkan kolom *clean_text* merupakan hasil normalisasi dan pembersihan teks melalui proses *case folding*, penghapusan karakter non-informatif, serta penyederhanaan struktur kata. Proses pra-pemrosesan ini bertujuan untuk mengurangi noise dan menghasilkan teks yang lebih terstruktur, sehingga siap digunakan pada tahap ekstraksi fitur dan klasifikasi sentimen.

```
...
Contoh teks setelah preprocessing:

                                full_text \
0 RUU BUMN Disahkan Selasa Depan Ada Danantara d...
1 Dengan adanya BPI Danantara isu2 kesulitan pem...
2 Danantara harus bebas dari titip-titipan ! Apa...
3 @AcruqAx1 Penempatan direktur utama definitif ...
4 RUU BUMN Segera Disahkan DPR Danantara Kapan M...
5 NASIB DANANTARA Ke Depan Jadi Tanda Tanya Besa...
6 Nasib Danantara ke depan jadi tanda tanya besa...
7 Badan Pengelola (BP) Investasi Daya Anagata Nu...
8 Payung Hukum Dikebut Pemerintah Mau Danantara ...
9 Ini episode seru. Pandu Syahrir putra Bang Cii...

                                clean_text
0 ruu bumn disahkan selasa depan ada danantara d...
1 dengan adanya bpi danantara isu kesulitan pemb...
2 danantara harus bebas dari titip titipan apa l...
3 penempatan direktur utama definitif pal di bpi...
4 ruu bumn segera disahkan dpr danantara kapan m...
5 nasib danantara ke depan jadi tanda tanya besa...
6 nasib danantara ke depan jadi tanda tanya besa...
7 badan pengelola bp investasi daya anagata nusa...
8 payung hukum dikebut pemerintah mau danantara ...
9 ini episode seru pandu syahrir putra bang ciil...
```

Gambar 4. Pra-pemrosesan Teks

3.3 Klasifikasi Utama Menggunakan IndoRoBERTa

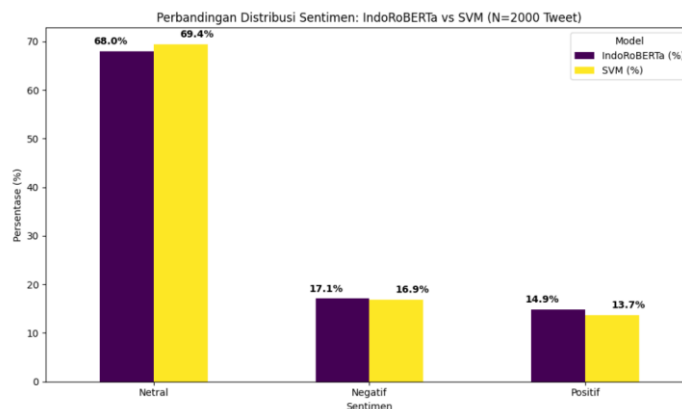
Setelah proses pra-pemrosesan, data teks diklasifikasikan menggunakan model IndoRoBERTa *base sentiment classifier* sebagai pengklasifikasi utama. Model berbasis transformer ini digunakan untuk menghasilkan prediksi sentimen awal karena kemampuannya dalam menangkap konteks bahasa Indonesia secara lebih baik. Hasil prediksi dari IndoRoBERTa memberikan gambaran awal mengenai kecenderungan sentimen pada data yang dianalisis dan menjadi dasar untuk tahap selanjutnya dalam sistem. Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.

3.4 Pseudo Labeling

Penerapan *pseudo-labeling* dengan memanfaatkan prediksi sentimen dari model IndoRoBERTa. Data yang tidak memiliki label manual tetapi memiliki nilai kepercayaan prediksi di atas ambang batas tertentu digunakan sebagai *pseudo-label*. Pendekatan ini memungkinkan pemanfaatan data tidak berlabel secara lebih optimal untuk memperluas data latih, sehingga variasi pola sentimen yang dipelajari model pada tahap pelatihan SVM menjadi lebih beragam.

3.5 Pelatihan dan Prediksi SVM

Pada gambar 5. menampilkan perbandingan distribusi sentimen antara hasil klasifikasi menggunakan model IndoRoBERTa dan model *Support Vector Machine* (SVM) pada 2.000 data tweet. Berdasarkan hasil yang ditampilkan, kedua model menunjukkan pola distribusi sentimen yang relatif serupa, dengan dominasi sentimen netral pada isu Danantara. Model SVM menghasilkan persentase sentimen netral sebesar 69,4%, sedikit lebih tinggi dibandingkan IndoRoBERTa sebesar 68,0%, sementara distribusi sentimen negatif dan positif menunjukkan perbedaan yang relatif kecil. Hal ini menunjukkan bahwa model SVM yang dilatih menggunakan representasi fitur TF-IDF dan memanfaatkan pseudo-label dari IndoRoBERTa mampu menangkap pola sentimen yang sejalan dengan model berbasis transformer. Kesamaan distribusi ini mengindikasikan bahwa pendekatan hybrid yang mengombinasikan IndoRoBERTa dan SVM efektif dalam menghasilkan klasifikasi sentimen yang konsisten pada data media sosial.



Gambar 5. Perbandingan Prediksi IndoRoBERTa vs SVM

3.6 Analisis Hasil

Tahap analisis hasil dilakukan dengan mengevaluasi distribusi sentimen yang dihasilkan oleh model SVM serta membandingkannya dengan distribusi sentimen dari model IndoRoBERTa. Analisis ini menunjukkan bahwa secara umum kedua model menghasilkan pola sentimen yang sejalan, meskipun terdapat perbedaan prediksi pada beberapa data. Perbedaan tersebut umumnya ditemukan pada teks yang bersifat ambigu atau memiliki konteks implisit, di mana model berbasis transformer cenderung lebih sensitif terhadap konteks kalimat dibandingkan model berbasis fitur seperti SVM.

3.7 Output dan Penyimpanan Hasil

Pada gambar 6 merupakan hasil tahap akhir dari sistem berupa penyajian dan penyimpanan hasil klasifikasi sentimen. Output yang dihasilkan berupa label sentimen untuk setiap data teks serta ringkasan distribusi sentimen terhadap isu Danantara. Hasil ini disimpan untuk keperluan analisis lanjutan dan dokumentasi penelitian, sehingga sistem yang dikembangkan dapat direproduksi dan digunakan kembali pada penelitian sejenis.

```
Jumlah tweet di mana RoBERTa dan SVM berbeda prediksi: 51 (2.4%)

Contoh 5 tweet dengan prediksi berbeda:

81. Teks: Jangan salah! Danantara hadir bukan untuk menyelamatkan BUMN yang sakit tapi untuk r
RoBERTa: Positif (conf: 0.94)
SVM: Netral (conf: 0.58)

84. Teks: @indoinsidenews Harusnya Danantara itu berbentuk perusahaan PT Danantara
RoBERTa: Positif (conf: 0.75)
SVM: Netral (conf: 0.60)

104. Teks: Asa BUMN Terbang Tinggi 2025 Bagaimana Danantara bisa menjadi superholding company
RoBERTa: Positif (conf: 0.88)
SVM: Netral (conf: 0.77)

109. Teks: @kegblgnunfaedh Keuntungan danantara katanya da meningkat btw afh iya
RoBERTa: Positif (conf: 0.78)
SVM: Netral (conf: 0.67)

128. Teks: @DS_yantie Pejabat nakal dibina nakalnya masukin ke danantara....
RoBERTa: Negatif (conf: 1.00)
SVM: Netral (conf: 0.79)
```

Gambar 6. Hasil Akhir

4. SIMPULAN

Penelitian ini berhasil mengembangkan sistem analisis sentimen media sosial berbahasa Indonesia terhadap isu Danantara dengan mengombinasikan model IndoRoBERTa sebagai pengklasifikasi utama dan *Support Vector Machine* (SVM) berbasis TF-IDF melalui pendekatan pseudo-labeling. Hasil penelitian menunjukkan bahwa model SVM yang dilatih menggunakan data berlabel hasil prediksi IndoRoBERTa mampu menghasilkan distribusi sentimen yang konsisten, dengan dominasi sentimen netral, serta menunjukkan kemampuan yang sejalan dengan model berbasis transformer. Pendekatan pseudo-labeling terbukti efektif dalam memperluas data latih tanpa ketergantungan pada pelabelan manual, sehingga meningkatkan efisiensi dan kelayakan sistem untuk penelitian akademik dengan sumber daya komputasi terbatas. Kelebihan sistem ini terletak pada desainnya yang efisien, mudah direproduksi, dan adaptif terhadap data teks tidak terstruktur, sementara

keterbatasannya masih bergantung pada kualitas pseudo-label dan belum dilakukan evaluasi performa kuantitatif secara menyeluruh.

5. SARAN

Penelitian selanjutnya disarankan untuk menambahkan evaluasi kuantitatif berbasis data berlabel manual dengan menggunakan metrik klasifikasi seperti accuracy, precision, recall, dan F1-score guna mengukur performa model secara objektif. Selain itu, perlu dilakukan eksperimen terhadap variasi ambang batas confidence pada proses *pseudo-labeling* untuk menganalisis pengaruh kualitas label semu terhadap kinerja model SVM. Optimalisasi parameter TF-IDF, seperti rentang *n-gram*, nilai *min_df* dan *max_df*, serta eksplorasi kernel dan parameter regularisasi pada SVM juga direkomendasikan untuk meningkatkan generalisasi model. Pengembangan lebih lanjut dapat mencakup fine-tuning model IndoRoBERTa pada data domain-spesifik guna memperbaiki kemampuan model dalam menangkap konteks sentimen yang lebih kompleks.

6. DAFTAR PUSTAKA

- [1] S. F. Yulianti, T. Rahmawati, P. Tawainella, N. M. Rantisi, and H. Mutiara Winata, “ANALISIS SENTIMEN KOMENTAR INSTAGRAM TERHADAP KASUS TUNTUTAN 17+8 MENGGUNAKAN ALGORITMA INDOBERT: KUALITAS KEPERCAYAAN PUBLIK,” 2025. [Online]. Available: <https://jurnal.panengeninsani.com/index.php/djods>
- [2] A. Kurniawan Santoso, “ANALISIS SENTIMEN TWITTER BAHASA INDONESIA MENGGUNAKAN PENDEKATAN MACHINE LEARNING,” *JIK*, vol. 6, no. 2, 2022.
- [3] D. Aby Vonega, A. Fadila, and D. Ely Kurniawan, “Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024,” 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [4] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.
- [5] F. Bei and S. Saepudin, “ANALISIS SENTIMEN APLIKASI TIKET ONLINE DI PLAY STORE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM),” 2021.
- [6] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” Oct. 2020, [Online]. Available: <http://arxiv.org/abs/2009.05387>
- [7] Agussabeth Adelina Pasaribu *et al.*, “Analisis Pendekatan Deep Learning Sebagai Upaya Endorsment Pendidikan,” *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, vol. 4, no. 2, pp. 9326–9333, Nov. 2025, doi: 10.31004/jerkin.v4i2.3120.
- [8] Z. Alhaq, A. Mustopa, and J. D. Santoso, “PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER.”
- [9] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, “Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma Svm,” 2021.
- [10] U. N. Jakarta, V. B. Lestari, and C. A. Hutagalung, “J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes,” pp. 22–2025, doi: 10.21009/j.
- [11] M. I. Abidin and E. W. Pamungkas, “ANALISIS SENTIMEN TERHADAP TIMNAS INDONESIA DI PIALA ASIA 2023 DENGAN MODEL TRANSFORMER BERBAHASA INDONESIA,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 482–496, Jul. 2025, doi: 10.36341/rabit.v10i2.6142.
- [12] L. Yang, B. Huang, S. Guo, Y. Lin, and T. Zhao, “A Small-Sample Text Classification Model Based on Pseudo-Label Fusion Clustering Algorithm,” *Applied Sciences (Switzerland)*, vol. 13, no. 8, Apr. 2023, doi: 10.3390/app13084716.
- [13] L. Afuan and N. Hidayat, “Sentiment Analysis of the Kampus Merdeka Program on Twitter Using Support Vector Machine and a Feature Extraction Comparison: TF-IDF vs. FastText,” *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1738–1753, Dec. 2024, doi: 10.47738/jads.v5i4.436.
- [14] U. N. Jakarta, V. B. Lestari, and C. A. Hutagalung, “J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes,” pp. 22–2025, doi: 10.21009/j.

- [15] F. Hibban Hilmi and A. Dwi Indriyanti, “Journal of Emerging Information System and Business Intelligence Sentiment Analysis of 2024 Election Fraud Using SVM and Naïve Bayes Algorithms,” *Journal of Emerging Information System and Business Intelligence*, vol. 5, no. 4, pp. 353–364, 2025.