

PERBANDINGAN METODE VEKTORISASI PADA ANALISA SENTIMENT, STUDI KASUS : CYBERBULLYING PADA KOMENTAR INSTAGRAM

I Putu Ramayasa¹⁾ I Gusti Ayu Desi Saryanti²⁾ I Komang Dharmendra³⁾ Edwar⁴⁾

Program Studi Sistem Informasi¹⁾²⁾³⁾ Program Studi Manajemen Informatika⁴⁾

Fakultas Informatika dan Komputer¹⁾²⁾³⁾, Fakultas Bisnis dan Fokasi⁴⁾

Institut Teknologi dan Bisnis STIKOM Bali¹⁾²⁾³⁾⁴⁾

ramayasa@stikom-bali.ac.id¹⁾, desi@stikom-bali.ac.id²⁾, dharmendra@stikom-bali.ac.id³⁾,

edwar.ridwan@stikom-bali.ac.id⁴⁾

ABSTRACT

Social media, as a global communication tool, has brought forth positive impacts but has also unveiled its darker facets, such as the rise in cyberbullying cases. This research proposes an approach to address this issue through sentiment analysis, with an emphasis on vectorization techniques. This approach combines three main vectorization techniques: Long Short-Term Memory (LSTM), Word2Vec, and Term Frequency-Inverse Document Frequency (TF-IDF), with three primary classification models: Random Forest, Naive Bayes, and Gradient Boosting. Research findings demonstrate that the utilization of LSTM in conjunction with the Random Forest algorithm yields the highest accuracy, reaching 92.5%, showcasing its ability to recognize intricate sentiment patterns. While Naive Bayes exhibits lower accuracy, Word2Vec significantly enhances sentiment pattern recognition within this algorithm. Employing Word2Vec also yields consistent outcomes in the Gradient Boosting algorithm, achieving approximately 86% accuracy, affirming its proficiency in identifying word relations in sentiment analysis concerning cyberbullying. Conversely, the TF-IDF vectorization technique delivers impressive results, achieving the highest accuracy of 96.25% in the Random Forest algorithm, adept at identifying influential keywords affecting sentiment. In conclusion, this research delineates that various vectorization methods wield a substantial impact on sentiment analysis of Instagram comments, particularly in the realm of cyberbullying detection. Therefore, the appropriate selection of vectorization methods holds paramount importance in crafting an effective sentiment analysis algorithm.

Keywords: Cyberbullying, Sentiment Analysis, Vectorization

ABSTRAK

Media sosial sebagai alat komunikasi global telah membawa dampak positif namun juga muncul sisi gelapnya, seperti peningkatan kasus *cyberbullying*. Penelitian ini mengajukan pendekatan melalui analisis sentimen, dengan penekanan pada teknik vektorisasi. Pendekatan ini menggabungkan tiga teknik vektorisasi utama, yaitu LSTM, Word2Vec, dan TF-IDF, dengan tiga model klasifikasi utama: *Random Forest*, *Naive Bayes*, dan *Gradient Boosting*. Hasil penelitian menunjukkan bahwa penggunaan LSTM dalam kaitannya dengan algoritma *Random Forest* memunculkan akurasi tertinggi, mencapai 92.5%, dengan kemampuan mengenali pola sentimen yang kompleks. Sementara *Naive Bayes* memiliki akurasi yang lebih rendah, Word2Vec memberikan peningkatan signifikan dalam pengenalan pola sentimen pada algoritma ini. Penggunaan Word2Vec juga menghasilkan hasil yang konsisten pada algoritma *Gradient Boosting* dengan akurasi sekitar 86%, menegaskan kemampuannya dalam mengidentifikasi relasi kata dalam analisis sentimen terkait *cyberbullying*. Di sisi lain, teknik vektorisasi TF-IDF memberikan hasil yang mengesankan, mencapai akurasi tertinggi 96.25% pada algoritma *Random Forest*, dengan kemampuan mengenali kata-kata kunci yang mempengaruhi sentimen. Kesimpulannya, penelitian ini menggambarkan bahwa berbagai metode vektorisasi memiliki dampak yang signifikan dalam analisis sentimen terhadap komentar Instagram, terutama dalam konteks deteksi *cyberbullying*. Oleh karena itu, pemilihan metode vektorisasi yang sesuai sangat penting dalam mengembangkan algoritma analisis sentimen yang efektif.

Kata Kunci : *Cyberbullying*, Analisa Sentimen, Vektorisasi

PENDAHULUAN

Media sosial menjadi salah satu alat komunikasi paling populer dan dapat diakses oleh siapa saja di seluruh dunia. Namun, meskipun menyediakan platform untuk berinteraksi dan berbagi informasi dengan mudah, media sosial juga memiliki sisi gelapnya, yaitu *Cyberbullying* atau perundungan di media sosial yang semakin sering terjadi akhir-akhir ini, terutama pada platform media sosial yang sangat populer seperti Instagram. Hal ini dapat mengancam kesehatan mental seseorang, meningkatkan tingkat depresi dan kecemasan, serta dapat memicu perasaan takut dan marah pada korban. Oleh karena itu, penting untuk melakukan analisis sentimen pada komentar di Instagram untuk mengidentifikasi apakah ada kecenderungan untuk terjadinya *cyberbullying* pada komentar tersebut.

Untuk mengatasi masalah ini, analisa sentimen dapat digunakan untuk mengidentifikasi dan mengklasifikasikan konten yang memiliki konten negatif atau positif[1]. Dalam melakukan analisis sentimen, salah satu faktor penting yang harus diperhatikan adalah teknik vektorisasi.

Vektorisasi pada sentiment analysis adalah suatu proses mengubah kata-kata atau teks menjadi representasi vektor numerik dalam ruang multi-dimensi[2]. Hal ini dilakukan untuk memungkinkan mesin pembelajaran mencari pola dan mengklasifikasikan sentimen dari teks yang diberikan. Dalam analisis sentimen, vektorisasi sangat penting karena memungkinkan mesin pembelajaran untuk memahami makna dari teks yang diberikan, bahkan jika kata-kata dalam teks itu sendiri tidak bermakna. Dalam konteks ini, vektorisasi dapat memperkuat kemampuan mesin pembelajaran untuk memahami bahasa manusia dengan lebih baik dan lebih akurat. Teknik ini bertujuan untuk mengubah teks menjadi bentuk numerik yang dapat dipahami oleh model mesin. Beberapa teknik vektorisasi yang populer digunakan adalah TF-IDF, LSTM, dan Word2Vec.

Namun, belum banyak penelitian yang membandingkan ketiga teknik vektorisasi

tersebut dalam melakukan analisis sentimen pada komentar Instagram, khususnya terkait dengan *cyberbullying*. Oleh karena itu, penelitian ini bertujuan untuk membandingkan ketiga teknik vektorisasi tersebut dalam melakukan analisis sentimen pada Instagram. Selain itu, penelitian ini juga akan membangun 3 model machine learning yaitu *Random Forest*, *Decision Tree*, dan *Gradient Boosting* untuk menguji kinerja ketiga teknik vektorisasi dalam memprediksi *cyberbullying* pada komentar Instagram.

TINJAUAN PUSTAKA INSTAGRAM

Instagram adalah salah satu platform media sosial yang paling populer di dunia, dengan lebih dari satu miliar pengguna aktif setiap bulannya. Instagram awalnya diluncurkan pada tahun 2010 sebagai aplikasi berbagi foto, tetapi sejak saat itu telah berkembang menjadi platform multimedia yang lebih luas dengan berbagai fitur seperti Instagram *Stories*, IGTV, dan *Reels*[3].

CYBERBULLYING

Cyberbullying atau penindasan melalui media digital merupakan sebuah bentuk kekerasan yang dilakukan dengan menggunakan teknologi informasi dan komunikasi, terutama internet dan media sosial. Dalam beberapa tahun terakhir, fenomena *cyberbullying* telah menjadi semakin umum dan telah menjadi perhatian masyarakat, terutama di kalangan remaja dan anak muda[4].

Cyberbullying juga dapat memiliki dampak jangka panjang pada korban, seperti hilangnya rasa percaya diri, menurunnya kinerja akademik, dan bahkan berpotensi mengalami tindakan kekerasan fisik. Oleh karena itu, penting untuk melakukan penelitian dan pengembangan metode untuk mencegah dan mengatasi *cyberbullying*.

Analisa sentimen

Analisa sentimen adalah salah satu teknik dalam bidang pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi dan mengekstrak sentimen dari teks, baik itu positif,

negatif, atau netral[5]. Dalam beberapa tahun terakhir, analisa sentimen telah menjadi topik yang semakin populer dalam dunia teknologi informasi, terutama dalam bidang social media monitoring dan customer feedback analysis.

VEKTORISASI

Vektorisasi pada machine learning adalah proses perubahan data teks menjadi representasi numerik yang dapat diolah oleh algoritma *machine learning*[6]. Representasi numerik ini disebut sebagai vektor, yang terdiri dari kumpulan bilangan riil atau bilangan bulat. Dalam konteks analisa sentiment, vektorisasi digunakan untuk mengubah teks pada data training menjadi representasi numerik yang dapat diproses oleh algoritma machine learning untuk melakukan klasifikasi pada data uji. Hal ini memungkinkan kita untuk mengetahui sentimen positif, negatif, atau netral dari sebuah teks.

TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode vektorisasi dokumen yang digunakan untuk mengukur seberapa penting suatu kata dalam sebuah dokumen atau korpus[6]. Metode ini sering digunakan dalam aplikasi NLP (*Natural Language Processing*) seperti klasifikasi dokumen, pencarian informasi, dan analisis sentimen[7].

TF-IDF terdiri dari dua konsep, yaitu *term frequency* dan *inverse document frequency*. *Term frequency* mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen, sedangkan *inverse document frequency* mengukur seberapa umum kata tersebut dalam seluruh korpus dokumen. Persamaan 1 menunjukkan persamaan untuk proses TF-IDF

$$TF - IDF = \frac{TF}{IDF} \quad (1)$$

LSTM

LSTM (*Long Short-Term Memory*) adalah salah satu jenis arsitektur model *Deep Learning* yang populer digunakan dalam tugas *Natural Language Processing* (NLP) seperti *Speech Recognition*, *Text Classification*, dan *Language Translation*. LSTM pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun

1997 sebagai solusi dari masalah menghilangnya *gradien* pada model RNN (*Recurrent Neural Network*)[2].

LSTM mengatasi masalah *vanishing gradient* dengan menggunakan mekanisme *gates* untuk mengatur aliran informasi. LSTM memiliki tiga *gates*, yaitu *forget gate*, *input gate*, dan *output gate*. *Forget gate* digunakan untuk memutuskan informasi mana yang harus dihapus dari *state* sebelumnya, *input gate* digunakan untuk memutuskan informasi mana yang harus ditambahkan ke dalam *state* saat ini, dan *output gate* digunakan untuk memutuskan informasi mana yang harus dikeluarkan pada *time step* saat ini[8].

LSTM telah berhasil digunakan dalam berbagai tugas NLP seperti *Text Classification*, *Sentiment Analysis*, dan *Language Translation*. Kelebihan LSTM adalah mampu mengatasi masalah *vanishing gradient* pada RNN dan mampu mempelajari hubungan jarak jauh dalam *sequence*. Namun, kelemahannya adalah kompleksitas model yang tinggi, sehingga membutuhkan waktu dan sumber daya komputasi yang besar dalam pelatihan model.

Word2Vec

Word2Vec adalah salah satu teknik populer dalam *Natural Language Processing* (NLP) untuk menghasilkan representasi vektor yang padat dari kata-kata dalam sebuah teks. Teknik ini diperkenalkan oleh Tomas Mikolov dari Google pada tahun 2013[9].

Word2Vec adalah model jaringan saraf tiruan yang digunakan untuk mempelajari representasi vektor dari kata-kata dalam sebuah teks[10][11]. Model ini bekerja dengan cara memproses sebuah teks dan menghasilkan representasi vektor yang padat untuk setiap kata dalam teks tersebut. Representasi vektor ini digunakan untuk mengukur kemiripan antara kata-kata dalam teks.

Word2Vec menggunakan dua arsitektur utama, yaitu *Continuous Bag-of-Words* (CBOW) dan *skip-gram*. CBOW memprediksi sebuah kata berdasarkan konteks kata-kata di sekitarnya, sedangkan *skip-gram* memprediksi konteks kata-kata di sekitar sebuah kata tertentu. Kedua arsitektur ini dapat menghasilkan representasi vektor yang padat untuk setiap kata dalam sebuah teks.

Word2Vec telah terbukti berhasil dalam berbagai tugas NLP, seperti pemrosesan bahasa alami, klasifikasi teks, dan analisis sentimen. Salah satu keunggulan Word2Vec adalah kemampuannya untuk menangkap makna kata-kata dan hubungan antara kata-kata, sehingga dapat menghasilkan representasi vektor yang lebih kaya dan informatif.

Random Forest

Random Forest adalah salah satu algoritma *Machine Learning* yang banyak digunakan dalam klasifikasi dan regresi. Algoritma ini digunakan untuk membangun model prediktif dengan menggabungkan beberapa pohon keputusan (*decision tree*) untuk meningkatkan akurasi dan meminimalkan *overfitting*[12].

Setiap pohon keputusan pada *random forest* dibangun secara acak dari subset data training dan variabel fitur. Kemudian, hasil dari setiap pohon dipilih sebagai output prediksi dengan melakukan voting atau *averaging*. Proses ini membantu mengurangi bias dan varian dari model, sehingga meningkatkan kinerja model pada data yang tidak dilihat sebelumnya.

Random Forest memiliki beberapa kelebihan dibandingkan dengan algoritma *machine learning* lainnya, seperti kemampuan untuk menangani data yang tidak seimbang, data yang memiliki banyak fitur, serta data yang memiliki *noise* atau *missing value*. Selain itu, *Random Forest* juga mudah digunakan dan dapat digunakan pada berbagai jenis masalah, seperti klasifikasi, regresi, dan *clustering*[12].

Namun, *Random Forest* juga memiliki beberapa kelemahan, seperti waktu komputasi yang relatif lama pada dataset yang besar, serta interpretasi yang sulit pada model yang kompleks. Oleh karena itu, perlu dilakukan penyesuaian parameter dan pengujian model untuk memastikan bahwa model *Random Forest* yang dibangun memiliki kinerja yang optimal pada data yang digunakan.

Decision tree

Decision tree adalah salah satu metode dalam *machine learning* yang digunakan untuk memecahkan masalah klasifikasi dan regresi. Konsep dasar dari *decision tree* adalah membuat sebuah struktur pohon keputusan

berdasarkan aturan-aturan yang diambil dari data training[13].

Dalam *decision tree*, setiap simpul (*node*) pada pohon keputusan merepresentasikan sebuah pengambilan keputusan berdasarkan fitur atau variabel pada data. Kemudian setiap cabang dari simpul tersebut akan merepresentasikan sebuah nilai atau kriteria dari fitur atau variabel tersebut. Proses ini dilakukan secara berulang hingga mencapai simpul terminal atau daun (*leaf node*) yang merepresentasikan kelas atau hasil akhir dari pengambilan keputusan.

Kelebihan dari *decision tree* adalah mudah dipahami dan diinterpretasikan, mampu menangani data yang tidak terstruktur dan dapat digunakan pada data kategori maupun numerik. Namun, *decision tree* juga memiliki kekurangan yaitu rentan terhadap *overfitting* dan tidak dapat menangani hubungan *non-linear* antar variabel[14].

Gradient Boosting

Gradient Boosting adalah metode pembelajaran mesin yang populer untuk masalah regresi dan klasifikasi. Konsep dasar *Gradient Boosting* adalah menggabungkan beberapa model pembelajaran mesin yang relatif lemah untuk membuat model yang lebih kuat[15].

Pada dasarnya, *Gradient Boosting* adalah jenis pohon keputusan yang ditingkatkan. Dalam pohon keputusan, data dianalisis berdasarkan serangkaian keputusan ya/tidak yang diambil dari fitur-fitur yang diberikan. *Gradient Boosting* memperbaiki model pohon keputusan dengan menggabungkan beberapa model dalam urutan berurutan. Prosesnya dimulai dengan membuat model pohon keputusan yang sederhana dan memperbaikinya dengan menambahkan model pohon keputusan berikutnya yang memfokuskan pada data yang kurang tepat diprediksi oleh model sebelumnya.

Salah satu kelebihan dari *Gradient Boosting* adalah kemampuannya untuk mengatasi masalah *overfitting* dengan lebih baik daripada model lain seperti *Random Forest*. Ini terjadi karena *Gradient Boosting* mempelajari dari kesalahan model sebelumnya dan mencoba memperbaikinya pada setiap langkah.

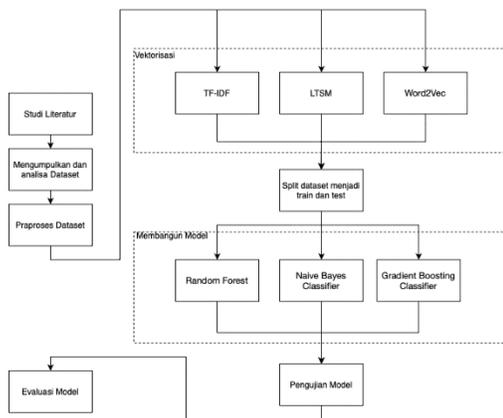
Selain itu, *Gradient Boosting* juga memungkinkan adanya berbagai pengaturan

hyperparameter yang dapat diatur untuk meningkatkan kinerja model. Namun, karena sifat kompleksitas model, Gradient Boosting dapat memerlukan waktu dan sumber daya yang lebih banyak untuk melatih dan mengevaluasi model[16].

METODE PENELITIAN

Kerangka Berpikir

Penelitian dilakukan dengan menggunakan data dari penelitian yang telah dilakukan oleh Wanda Athira Luqyana, Imam Cholissodin, dan Rizal Setya Perdana yang berasal dari Fakultas Ilmu Komputer, Universitas Brawijaya[17]. Dengan dataset adalah opini yang dibagikan secara publik dengan jumlah 400 data yang terbagi menjadi dua kelas, 200 data kelas positif dan 200 data kelas negatif, dan dataset berisi opini yang dipublikasikan secara publik pada platform Instagram.



Gambar 1. Kerangka Berpikir

Metode Penelitian

Dataset dipersiapkan melalui serangkaian langkah praproses yang penting. Pertama-tama, semua karakter dalam dataset diubah menjadi huruf kecil guna memastikan konsistensi dalam analisis. Selanjutnya, karakter non-abad seperti angka dan simbol khusus dihilangkan dari teks untuk membersihkan data dari unsur yang tidak relevan. Praproses juga mencakup penghapusan emotikon yang dapat memengaruhi interpretasi teks secara tidak langsung. Tautan atau hyperlink yang mungkin ada dalam teks juga dihapus agar fokus utama

tetap pada konten teks itu sendiri. Selanjutnya, tanda mention seperti @username dihilangkan agar analisis tidak terganggu oleh informasi pengguna tertentu.

Satu tahap penting dalam praproses adalah penghapusan kata-kata umum atau "stopword." Kata-kata ini sering muncul dalam teks tetapi kurang memiliki makna khusus dalam konteks analisis. Dengan menghapus stopwords, teks yang tersisa menjadi lebih padat dan fokus pada informasi yang lebih bermakna. Terakhir, untuk memperoleh bentuk kata dasar, langkah stemming dilakukan menggunakan model Sastrawi. Proses ini membantu mengatasi variasi kata yang memiliki akar kata yang sama, sehingga analisis lebih akurat dan komprehensif. Dengan melakukan langkah-langkah praproses, dataset menjadi lebih siap untuk dianalisis secara lebih mendalam dan dapat diandalkan untuk mendapatkan wawasan yang berarti dari konten teks yang disajikan.

Proses dilanjutkan dengan melakukan vektorisasi, dalam pembangunan model machine learning, vektorisasi mengacu pada proses mengubah data dalam bentuk non-numerik menjadi representasi numerik (vektor) sehingga dapat dimanfaatkan oleh algoritma machine learning. Data dalam bentuk teks, gambar, audio, atau bentuk non-numerik lainnya sulit untuk langsung dimengerti oleh algoritma machine learning yang umumnya bekerja dengan angka.

Vektorisasi memungkinkan data yang kompleks atau tidak terstruktur seperti teks atau gambar diubah menjadi format yang bisa diolah oleh algoritma machine learning, seperti algoritma klasifikasi atau regresi. Dalam konteks data teks, vektorisasi mengubah setiap dokumen atau kalimat menjadi vektor numerik dengan menggunakan metode tertentu.

Pada penelitian ini dilakukan proses vektorisasi dengan menggunakan 3 metode, dimana dataset akan dilakukan vektorisasi dengan menggunakan setiap metode vektorisasi, TF-IDF, LTSM, dan Word2Vec yang akan dilakukan secara terpisah, kemudian data hasil vektorisasi dari masing-masing metode vektorisasi akan dibangun model menggunakan algoritma Random Forest, naive bayes classification, dan Gradient Boosting untuk mengetahui performa dari masing-masing

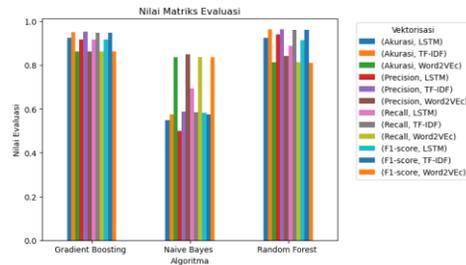
metode vektorisasi jika diterapkan pada algoritma tertentu.

Tabel 1. Hasil Pengujian

Algoritma	Akurasi	Precision	Recall	F1-score
LSTM				
Random Forest	0,925	0,941	0,889	0,914
Naive Bayes	0,550	0,500	0,694	0,581
Gradient Boosting	0,925	0,917	0,917	0,917
Word2Vec				
Random Forest	0,813	0,842	0,813	0,812
Naive Bayes	0,838	0,851	0,838	0,838
Gradient Boosting	0,863	0,862	0,863	0,862
TF-IDF				
Random Forest	0,963	0,963	0,961	0,962
Naive Bayes	0,575	0,589	0,586	0,574
Gradient Boosting	0,950	0,953	0,947	0,949

Penggunaan metode vektorisasi dalam analisis sentimen memiliki implikasi yang signifikan dalam memahami dampak *cyberbullying* dalam komentar Instagram. metode LSTM digunakan untuk menganalisis dampak penggunaan vektorisasi pada algoritma Random Forest Classifier, Naive Bayes Classifier, dan Gradient Boosting Classifier. Algoritma Random Forest Classifier mencapai akurasi sebesar 92.5%, dengan precision dan recall yang tinggi. Ini mengindikasikan bahwa metode LSTM efektif dalam mengenali pola sentimen yang rumit dalam komentar Instagram terkait *cyberbullying*. Namun, Naive Bayes Classifier menunjukkan hasil akurasi yang lebih rendah sekitar 55%, menunjukkan bahwa Naive Bayes kurang mampu menangani kompleksitas analisis sentimen dalam kasus ini. Selanjutnya, metode Word2Vec digunakan untuk mengembangkan vektorisasi pada algoritma Random Forest Classifier dan Naive Bayes Classifier. Dengan akurasi 81.25%, algoritma Random Forest Classifier memiliki hasil yang baik, menunjukkan kemampuan Word2Vec dalam menggambarkan konteks kalimat dalam analisis sentimen. Naive Bayes

Classifier juga menunjukkan peningkatan dengan akurasi 83.75%, namun dengan nilai recall yang sedikit lebih rendah. Hasil ini menyoroti pentingnya pendekatan Word2Vec dalam pengenalan pola sentimen dalam komentar Instagram.



Gambar 2. Grafik Hasil Evaluasi

Metode Word2Vec diterapkan pada algoritma klasifikasi lainnya, yaitu Gradient Boosting Classifier. Algoritma ini menghasilkan akurasi, precision, recall, dan F1-score sekitar 86%. Penggunaan Word2Vec berkontribusi pada hasil yang konsisten, menegaskan kemampuannya dalam mengidentifikasi dan menggambarkan relasi kata-kata dalam analisis sentimen yang terkait dengan *cyberbullying*. Metode vektorisasi TF-IDF diterapkan pada algoritma Random Forest, Naive Bayes, dan Gradient Boosting. Hasilnya mengesankan, dengan akurasi tertinggi mencapai 96.25% pada algoritma Random Forest. TF-IDF membantu algoritma dalam mengenali kata-kata kunci yang mempengaruhi sentimen dan memberikan hasil yang konsisten dan baik pada ketiga algoritma.

SIMPULAN

Metode vektorisasi memiliki peran sentral dalam analisis sentimen terkait *cyberbullying* pada komentar Instagram. Penggunaan metode LSTM, Word2Vec, dan TF-IDF membantu dalam mengenali pola sentimen, konteks kalimat, dan kata-kata kunci yang berkontribusi pada hasil evaluasi yang baik. Untuk pengembangan lebih lanjut, disarankan untuk menjajaki kombinasi metode vektorisasi yang lebih kompleks atau pendekatan deep learning untuk memperoleh wawasan yang lebih mendalam tentang dampak *cyberbullying* dalam ranah media sosial.

DAFTAR PUSTAKA

- [1] I. K. Dharmendra, I. G. N. A. Kusuma, I. A. M. C. Dewi, and Edwar, "IMPLEMENTASI TEXT MINING UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 3, Art. no. 3, Jul. 2023, Accessed: Aug. 18, 2023. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2504>
- [2] "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) | SpringerLink." <https://link.springer.com/article/10.1007/s11277-021-08580-3> (accessed Apr. 28, 2023).
- [3] K. Hammar, S. Jaradat, N. Dokoohaki, and M. Matskin, "Deep Text Mining of Instagram Data without Strong Supervision," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Dec. 2018, pp. 158–165. doi: 10.1109/WI.2018.00-94.
- [4] R. Bayari and A. Bensefia, "Text Mining Techniques for Cyberbullying Detection: State of the Art," *Adv. sci. technol. eng. syst. j.*, vol. 6, no. 1, pp. 783–790, Feb. 2021, doi: 10.25046/aj060187.
- [5] K. Dharmendra, K. O. Saputra, and I. N. Pramaita, "Analisa Sentiment Untuk Opini Alumni Perguruan Tinggi," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 2, pp. xxxx–xxxx, Jul. 2019, doi: 10.24843/mite.2019.v18i02.p11.
- [6] B. B. Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *INISTA (Journal of Informatics Information System Software Engineering and Applications)*, vol. 3, no. 2, Art. no. 2, Jun. 2021, doi: 10.20895/inista.v3i2.218.
- [7] M. P. Simatupang and D. P. Utomo, "ANALISA TESTIMONIAL DENGAN MENGGUNAKAN ALGORITMA TEXT MINING DAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCE (TF-IDF) PADA TOKO ALLMEEART," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Art. no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1697.
- [8] "Sentiment Analysis of Comment Texts Based on BiLSTM | IEEE Journals & Magazine | IEEE Xplore." <https://ieeexplore.ieee.org/abstract/document/8684825> (accessed Jul. 10, 2023).
- [9] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," *Procedia Computer Science*, vol. 179, pp. 728–735, Jan. 2021, doi: 10.1016/j.procs.2021.01.061.
- [10] S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, "Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," in *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, Jan. 2022, pp. 163–167. doi: 10.1109/ISMODE53584.2022.9742906.
- [11] "The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)." <http://jurnal.iaii.or.id/index.php/RESTI/article/view/3711> (accessed Aug. 25, 2023).

- [12] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm," *Procedia Computer Science*, vol. 161, pp. 765–772, Jan. 2019, doi: 10.1016/j.procs.2019.11.181.
- [13] Md. A. Rahman and Y. A. Akter, "Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naïve Bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–4. doi: 10.1109/ICASERT.2019.8934502.
- [14] Y. T. Samuel, J. J. Hutapea, and B. Jonathan, "Predicting the Timeliness of Student Graduation Using Decision Tree C4.5 Algorithm in Universitas Advent Indonesia," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Jul. 2019, pp. 276–280. doi: 10.1109/ICTS.2019.8850948.
- [15] H. Reddy, N. Raj, M. Gala, and A. Basava, "Text-mining-based Fake News Detection Using Ensemble Methods," *Int. J. Autom. Comput.*, vol. 17, no. 2, pp. 210–221, Apr. 2020, doi: 10.1007/s11633-019-1216-5.
- [16] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, "Sentiment Analysis for e-Payment Service Providers Using Evolutionary eXtreme Gradient Boosting," *IEEE Access*, vol. 8, pp. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.
- [17] "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer." <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051> (accessed Apr. 28, 2023).