

Analisis Perbandingan Metode Tf-Idf dan Word2vec pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal di Indonesia

Ivan Rifky Hendrawan^{*1}, Ema Utami², Anggit Dwi Hartanto³

^{1,2,3} Teknik Informatika, Universitas Amikom Yogyakarta

E-mail: ^{*}ivanrifky@students.amikom.ac.id, ²ema.u@amikom.ac.id, ³anggit@amikom.ac.id

Abstrak

Secara umum, proses menghasilkan setiap ulasan produk pada dasarnya terkait dengan tingkat rating, yang membuat pengguna memberikan komentar yang bias. Analisis sentimen dapat diterapkan pada ulasan produk Marketplace sehingga dapat digunakan sebagai saran perbaikan produk untuk penjual dan pesaing sehingga dapat mengetahui produk apa yang disenangi dan dibutuhkan oleh masyarakat. Penelitian ini menggunakan algoritma XGBoost dengan menggunakan dataset bahasa Indonesia yang dikombinasikan dengan TF-IDF dan Word2vec dan akan dievaluasi kombinasi mana yang lebih baik dalam mengklasifikasikan data teks yang tidak seimbang. Berdasarkan penelitian yang telah dilakukan, dua vector space TF-IDF dan Word2vec menghasilkan nilai F1-Score yang berbeda pada algoritma klasifikasi XGBoost, kombinasi Word2vec+XGboost menghasilkan nilai F1-Score lebih tinggi 0.941% dibanding TF IDF+XGBoost 0.940%. Hal ini dikarenakan word2vec lebih baik karena memiliki keunggulan dapat melihat hubungan semantik antar kata.

Kata kunci: word2vec, tfidf, sentimen analisis, XGBoost,

1. PENDAHULUAN

Tingkat penetrasi internet di Indonesia mencapai 73,7 persen dari total populasi pada awal tahun 2022 [1]. Pesatnya pertumbuhan pengguna internet menjadikan Indonesia sebagai pasar yang menjanjikan untuk marketplace, marketplace adalah sebuah aplikasi online yang memfasilitasi jual beli proses dari berbagai toko [2]. Marketplace menjadi media transaksi online populer dengan berbagai fitur yang dimilikinya, seperti mencari produk, melakukan transaksi ke memberikan ulasan produk. Secara umum, proses menghasilkan setiap ulasan produk pada dasarnya terkait dengan tingkat rating, yang membuat pengguna memberikan komentar yang bias[3].

Ulasan produk bisa berisi testimoni baik atau buruk dari pengguna yang telah membeli produk. Ulasan bisa digunakan sebagai sumber data untuk membuat berbagai keputusan manajemen [4]. Ulasan online sangat penting dalam mendukung pengambilan keputusan pembelian konsumen [5]. Salah satu teknik untuk mengetahui respon atau ulasan suatu produk adalah menggunakan analisis sentimen. Hal ini juga bisa digunakan untuk mempelajari pandangan individu, perilaku, perasaan terhadap orang lain, diri sendiri, masalah yang berkembang serta kegiatan yang selama ini sudah dilakukan[6].

Beberapa penelitian terkait sentiment analisis ulasan produk telah banyak dilakukan dengan berbagai metode dan hasil sentimen yang beragam. Seperti penelitian yang dilakukan oleh [2],[7] menggunakan algoritma naive Bayes dan KNN, metode ini juga pernah dilakukan untuk analisis sentiment review produk di aplikasi shopee[8] dan beberapa gabungan beberapa aplikasi ecommerce[9]. Selain naive bayes dan KNN, metode Support Vector Machine (SVM) juga pernah digunakan sebagai penunjang analisis sentimen pengambilan keputusan terhadap suatu produk[23] dan untuk menentukan klasifikasi pada ulasan produk dalam lima E-Commerce di Indonesia [10].

Menurut penelitian[7] Naive Bayes menghasilkan nilai akurasi lebih tinggi dibanding KNN dalam mengklasifikasikan ulasan produk di ecommerce. Jumlah data tentunya akan mempengaruhi pemilihan metode yang tepat. Mayoritas penelitian terdahulu hanya menggunakan

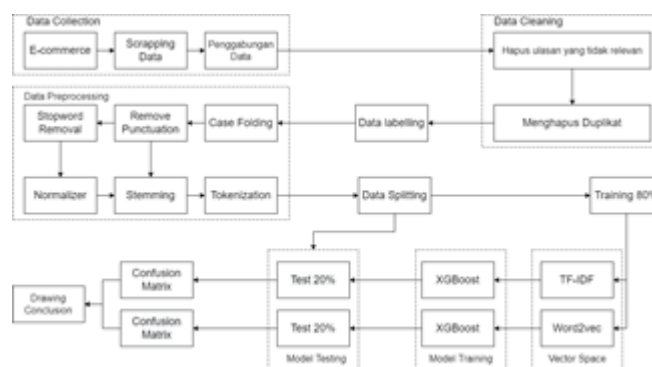
data yang relatif sedikit dan memiliki kelas data seimbang. Hal lain yang berpengaruh adalah keseimbangan data. Padahal tidak semuanya data yang ditemui dilapangan berjumlah seimbang.

Saat ini metode Xtreme gradient boost sangat populer digunakan untuk mengatasi ketidakseimbangan data. menurut [11] algoritma ini cukup baik dalam menangani data yang tidak seimbang. Hal ini didukung oleh Penelitian selanjutnya [9] membahas perbandingan kinerja TF-IDF pada dua algoritma, yaitu XGBoost dan Logistic Regression dengan menggunakan bahasa Bangla. Hasil penelitian ini membuktikan bahwa dari data pembelajaran mesin dengan TF-IDF XGBoost mencapai akurasi lebih tinggi dibandingkan dengan TF-IDF Logistic Regression.

Berdasarkan penelitian sebelumnya, algoritma XGBoost terbukti memiliki performa yang baik untuk melakukan sentimen analisis. Oleh sebab itu penelitian ini juga menggunakan algoritma XGBoost dengan menggunakan dataset bahasa Indonesia yang dikombinasikan dengan TF-IDF dan Word2vec dan akan dievaluasi kombinasi mana yang lebih baik dalam mengklasifikasikan data teks yang tidak seimbang.

2. METODE PENELITIAN

Penelitian ini memiliki 7 tahapan yaitu pengambilan data melalui metode scrapping, data splitting, labelling, cleaning data, preprocessing, tahap pelatihan model dan evaluasi. Metodologi yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Sedangkan detail dari setiap tahapan pada Gambar 1. dijelaskan pada sub bab A sampai dengan G.



Gambar 1. Metodologi Penelitian.

2.1. Data Collection

Pengumpulan review produk dilakukan dengan cara scrapping Shopee Marketplace dengan tool yang dibuat secara mandiri oleh penulis menggunakan Python. Scrapping dilakukan pada brand lokal terkenal dengan produk apparel terlaris. Kemudian data hasil scrapping akan digabungkan membentuk kumpulan data. Kemudian data yang sudah digabungkan itu akan disimpan dalam format .csv. Data yang digunakan adalah teks ulasan dan rating.

2.2. Data Splitting

Pembagian Data Data yang telah selesai di pre-processing akan dibagi menjadi dua bagian, yaitu 80% data train, 20% data test. Data train digunakan untuk melatih model XGBoost pada klasifikasi ulasan teks, sedangkan data test digunakan untuk memprediksi ulasan teks.

2.3. Data Labelling

Data yang telah dibersihkan kemudian diberikan label baik dan buruk. Kriteria buruk adalah untuk rating 1 sampai 3, sedangkan kriteria baik rating 4 sampai 5

2.4. Data Cleaning

Data yang telah dikumpulkan pada langkah sebelumnya dibersihkan terlebih dahulu agar lebih optimal, Ulasan yang spam atau yang tidak relevan akan dihapus agar proses sentimen kedepanya nanti bisa optimal. Ulasan yang redundant atau duplikat juga akan dihapus.

2.5. Data Pre-processing

Menurut[17] tahap preprocessing dimulai dari normalisasi, dimana kata atau frase dikembalikan ke kata baku. Proses selanjutnya adalah pembersihan data, dimana angka, tanda baca, dan tanda unik dihilangkan. Langkah selanjutnya adalah lower case, di mana semua kata distandarisasi dengan huruf kecil [18]. Setelah semua kata direduksi menjadi huruf kecil, masuk ke proses stopword removal. Proses ini merupakan proses menghilangkan konjungsi atau kata hubung.

Setelah proses penghilangan kata sambung, barulah proses stemming, dimana kata dikembalikan ke kata dasar[19]. Tahap terakhir dari proses preprocessing yang terakhir adalah tokenization, dimana sebuah kalimat atau paragraf akan diubah menjadi sebuah token dengan beberapa arti [20]. Untuk contoh kata yang diolah bias dilihat di table 1.

Tabel 1. Proses Text Preprocessing

N0	Text Preprocessing	
	Tahap	Sebelum Sesudah
1	Teks Kotor	Bahan nya halus adem nyaman dipakai ... bakalan order lagii nih 😊😊
2	Case Folding	bahan nya halus adem nyaman dipakai ... bakalan order lagii nih 😊😊
3	Remove Punctuation	bahan nya halus adem nyaman dipakai bakalan order lagii nih
4	Remove Number dan short word	bahan halus adem nyaman dipakai bakalan order lagii
5	Word Normalize	bahan halus adem nyaman dipakai bakalan order lagi
6	Stopword Removal	bahan halus adem nyaman dipakai bakalan order lagi
7	Tokenization	"bahan", "halus", "adem", "nyaman", "bakalan", "order", "lagi"
8	Stemming	"bahan", "halus", "adem", "nyaman", "bkal", "order", "lagi"

2.6. Tahap Pelatihan dengan XGBoost

Tahap pelatihan XGBoost Model yang akan digunakan untuk training klasifikasi adalah XGBoost. Dalam tahap pelatihan ini akan menggunakan dua skenario yang pertama adalah menggunakan TFIDF + XGBoost, dan Word2vec + XGBoost. Selanjutnya, model yang telah dilatih akan diuji di setiap skenario dengan menggunakan data yang test sebesar 20%.

2.7. Tahap Conclusion

Conclusion merupakan tahap untuk melakukan evaluasi performa (akurasi, presisi, recall dan f1 score) yang dihasilkan oleh model Algoritma XGBoost saat menggunakan TF-IDF dan Word2vec dataset yang sama. Dalam tahap ini akan menggunakan confusion matrix, confusion matrix adalah tabulasi dari [17] perhitungan yang didasari pada evaluasi kinerja model klasifikasi berdasarkan jumlah objek penelitian yang diprediksi dengan benar dan salah. Secara singkat confusion matrix memberikan perincian terkait kesalahan klasifikasi[18]. Dalam tahap ini akan dibandingkan model mana yang memiliki nilai F1-Score yang paling tinggi.

0.940% lebih baik dibandingkan dengan komposisi perbandingan 70.30 hanya menghasilkan 0.935% untuk TFIDF+XGBoost dan 0.939% untuk Word2vec+XGBoost. Dari dua kali scenario pengujian Word2vec menghasilkan nilai F1-Score lebih baik dibandingkan dengan TF-IDF. Hal ini dikarenakan kelebihan word2vec mampu memproses relasi semantik antar kata dibanding dengan TF-IDF.

Kombinasi Word2vec+XGBoost tentunya menghasilkan nilai F1-Score yang lebih baik dibandingkan penelitian sebelumnya yang dilakukan oleh [12] dengan menggunakan Word2vec+CNN yang hanya menghasilkan nilai 0.925%. Kombinasi TF-IDF+XGBoost pada penelitian ini juga menghasilkan nilai F1-Score lebih baik 0,940% dibandingkan penelitian sebelumnya yang dilakukan oleh [9] dengan metode yang sama hanya menghasilkan nilai 0.91%. Hal ini bisa berbeda karena pada penelitian saat ini menggunakan fitur normalizer kata pada tahap preprocessing.

4. KESIMPULAN

Analisis sentimen dapat diterapkan pada ulasan produk Marketplace sehingga dapat digunakan sebagai saran perbaikan produk untuk penjual dan pesaing sehingga dapat mengetahui produk apa yang disenangi dan dibutuhkan oleh masyarakat. Bagi pembeli, analisis sentimen review produk juga bisa menjadi pertimbangan sebelum memutuskan untuk membeli suatu produk karena review produk merupakan pendapat orang yang pernah membeli produk tersebut sebelumnya. Berdasarkan penelitian yang telah dilakukan, dua vector space TF-IDF dan Word2vec menghasilkan nilai F1-Score yang berbeda pada algoritma klasifikasi XGBoost, kombinasi Word2vec+XGboost menghasilkan nilai F1-Score lebih tinggi 0.941% dibanding TF-IDF+XGBoost 0.940%. Hal ini dikarenakan word2vec lebih baik karena memiliki keunggulan dapat melihat hubungan semantik antar kata.

Dalam penelitian selanjutnya, akan lebih baik untuk diterapkan deteksi kalimat sarkasme untuk menemukan ulasan yang berupa sindiran terhadap suatu produk sehingga dapat meningkatkan kualitas analisis sentimen.

DAFTAR PUSTAKA

- [1] Simon Kemp, 5 April 2022, Digital 2022 Indonesia :Internet use in Indonesia 2022, <https://datareportal.com/reports/digital-2022-indonesia?rq=indonesia%202022>.
- [2] N. Rohman, R. Luviana Musyarofah, E. Utami, and S. Raharjo, "Natural Language Processing on Marketplace Product Review Sentiment Analysis," in 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020, pp. 1–5, doi: 10.1109/ICORIS50180.2020.9320827.
- [3] X. Wang, T. Zhou, X. Wang, and Y. Fang, "Harshness-aware sentiment mining framework for product review," *Expert Systems with Applications*, vol. 187, p. 115887, 2022, doi: <https://doi.org/10.1016/j.eswa.2021.115887>.
- [4] J.-W. Bi, Y. Liu, and Z.-P. Fan, "Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking," *Information Sciences*, vol. 504, pp. 293–307, 2019, doi: <https://doi.org/10.1016/j.ins.2019.07.025>.
- [5] Q. Wang, W. Zhang, J. Li, F. Mai, and Z. Ma, "Effect of online review sentiment on product sales: The moderating role of review credibility perception," *Computers in Human Behavior*, vol. 133, p. 107272, 2022, doi: <https://doi.org/10.1016/j.chb.2022.107272>.
- [6] Kevin, V. et al. (2020) "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment

- Analysis Using Support Vector Machine Based on Particle Swarm Optimization),” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 9(2), hal. 162–170 FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
- [7] Y. Yennimar and R. Rizal, “Comparison of Machine Learning Classification Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency,” *Sinkron*, vol. 4, p. 268, 2019, doi: 10.33395/sinkron.v4i1.10416
- [8] Sihombing, L., Hannie, H. dan Dermawan, B. (2021) “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier,” *Edumatic: Jurnal Pendidikan Informatika*, 5, hal. 233–242. doi: 10.29408/edumatic.v5i2.4089
- [9] M. T. Akter, M. Begum, and R. Mustafa, “Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors,” in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 40–44, doi: 10.1109/ICICT4SD50815.2021.9396910
- [10] S. F. N. H. R. JAYADI, “Sentiment Analysis Of Indonesian E-Commerce Product Reviews Using Support Vector Machine Based Term Frequency Inverse Document,” vol. 99, no. 17, pp. 4316–4325, 2022
- [11] K. Afifah, I. N. Yulita, and I. Sarathan, “Sentiment Analysis on Telemedicine App Reviews using XGBoost Classifier,” in 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021, pp. 22–27, doi: 10.1109/ICAIBDA53487.2021.9689735.
- [12] A. Nurdin, B. Seno aji, A. Bustamin, and Z. Abidin, “PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS,” *Jurnal Tekno Kompak*, vol. 14, p. 74, 2020, doi: 10.33365/jtk.v14i2.732
- [13] Willy, Setiawan, E. dan Nugraha, F. (2019) “Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter,” in, hal. 114–119. doi: 10.1109/IC3INA48034.2019.8949601
- [14] Kurniawan, F. W. dan Maharani, W. (2020) “Analisis Sentimen Twitter Bahasa Indonesia dengan Word2Vec,” 7(2), hal. 7821–7829
- [15] Hermanto, D. T. et al. (2021) “Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online,” *Citec Journal*, Vol. 8, No, hal. 64–77
- [16] Mo, H. et al. (2019) “Developing window behavior models for residential buildings using XGBoost algorithm,” *Energy and Buildings*, 205, hal. 109564. doi: 10.1016/j.enbuild.2019.109564
- [17] Zhang, W., Li, Y. dan Wang, S. (2019) “Learning document representation via topic-enhanced LSTM model,” *Knowledge-Based Systems*, 174, hal. 194–204. doi: <https://doi.org/10.1016/j.knosys.2019.03.007>.
- [18] Kusumaningrum, R. et al. (2021) “Sentiment analysis of Indonesian hotel reviews: from classical machine learning to deep learning,” *International Journal of Advances in Intelligent Informatics*, 7, hal. 292. doi: 10.26555/ijain.v7i3.737.
- [19] Amin, S. et al. (2021) “Early Detection of Seasonal Outbreaks from Twitter Data Using Machine Learning Approaches,” *Complexity*, 2021, hal. 1–12. doi: 10.1155/2021/5520366.
- [20] Septian, J. A., Fahrudin, T. M. dan Nugroho, A. (2019) “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor,” *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, hal. 43–49.
- [21] Kotu, V. dan Deshpande, B. (2019) “Chapter 8 - Model Evaluation,” in Kotu, V. dan Deshpande, B. (ed.) *Data Science (Second Edition)*. Second Edition. Morgan Kaufmann, hal. 263–279. doi: <https://doi.org/10.1016/B978-0-12-814761-0.00008-3>.

- [22] Gorunescu, F. (2010). *Data Mining Concepts, Models and Techniques*. Springer International Publishing. <https://doi.org/10.1007/978-3-642-19721-5>
- [23] Basani, Y. et al. (2019) "Application of Sentiment Analysis on Product Review E-Commerce," *Journal of Physics: Conference Series*, 1175, hal. 12103. doi: 10.1088/1742-6596/1175/1/012103.