

Online : ejournal.stmikbinapatria.ac.id/index.php/DS/issue/ ISSN : 1978-5569

PERBANDINGAN ALGORITMA NAÏVE BAYES, SVM DAN XGBOOST DALAM KLASIFIKASI TEKS SENTIMEN MASYARAKAT TERHADAP PRODUK LOKAL DI INDONESIA

Ivan Rifky Hendrawan

“Teknik Informatika” Universitas Amikom Yogyakarta

Email : ivanrifky@students.amikom.ac.id

Abstract

Marketplace has become a shopping medium with promised facilities, one of the features that can be used for research is online reviews. Online reviews can be used as an increasing factor in the buying process as well as a data source for decision making because of the development of e-commerce, many fake reviews so that more and more consumers are worried about shopping online. This cannot be denied because from customer reviews it can be seen the level of loyalty and liking for what is purchased. Sentiment analysis can be applied as a method to find out people's responses on social media and also to find out product performance. Based on the research that has been done, that the combination of Word2vec + XGBoost produces a higher F1 score of 0.941 followed by TF-IDF + XGBoost 0.940. Meanwhile, the SVM algorithm using TF-IDF and Word2vec vector space only produces 0.938 and 0.939. Meanwhile, Naïve Bayes has an F1-Score of 0.915 with TF-IDF and 0.900 with word2vec. Classification with word2vec in representing words into vectors is better than TF-IDF, this is because word2vec is able to optimally utilize word similarity.

Keywords: Naïve Bayes, XGBoost, SVM, Sentiment Analysis, Review.

Abstrak

Marketplace sudah menjadi media belanja dengan fasilitas yang menjanjikan, salah satu fitur yang bisa dimanfaatkan untuk penelitian adalah ulasan online. Ulasan online dapat digunakan sebagai faktor penentuan dalam proses pembelian serta sebagai sumber data untuk landasan pengambil keputusan karena dengan berkembangnya e-commerce, semakin banyak ulasan palsu sehingga semakin banyak konsumen yang khawatir akan tertipu dalam belanja online. Hal ini tidak bisa dipungkiri dikarenakan dari review pelanggan dapat diketahui tingkat keloyalan dan kesukaan terhadap apa yang dibeli. Analisis sentimen dapat diterapkan sebagai metode untuk mengetahui respon orang di media social dan juga untuk mengetahui peforma produk. Berdasarkan penelitian yang telah dilakukan, bahwa kombinasi Word2vec + XGBoost menghasilkan F1 Score lebih tinggi yaitu sebesar 0.941 diikuti dengan TF-IDF + XGBoost 0.940. Sementara untuk algoritma SVM dengan menggunakan vector space TF-IDF dan Word2vec hanya menghasilkan 0.938 dan 0.939. Sedangkan untuk Naïve Bayes memiliki F1-Score 0.915 dengan TF-IDF dan 0.900 dengan word2vec. Klasifikasi dengan word2vec dalam merepresentasikan kata menjadi vektor lebih baik dibandingkan dengan TF-IDF, hal ini dikarenakan word2vec mampu mengolah kesamaan kata dengan optimal.

Kata kunci: Naïve Bayes, XGBoost, SVM, Analisis Sentimen, Ulasan

1. Pendahuluan

Indonesia saat ini tidak kekurangan karya, talenta serta produk-produk yang berkualitas (Kominfo.kulonprogo, 2022). idEa mengatakan ekonomi digital tahun 2021 ditaksir berjumlah sekitar Rp401 triliun, pertumbuhannya di atas 40 persen(Ekonomi.Bisnis, 2022). Penggunaan internet semakin meningkat di negara

Indonesia hingga mencapai maksimal 73,7 persen dari total populasi saat awal tahun 2022(S.Kemp., 2022) hal ini didukung oleh data dari kepios yang mencapai 2,1 juta (+1,0 persen) antara tahun 2021 sampai 2022(Simon Kemp, 2022). Pesatnya pertumbuhan pengguna internet menjadikan Indonesia sebagai pasar yang menjanjikan untuk marketplace(Rohman et al., 2020).

Marketplace menjadi media transaksi online populer dengan berbagai fitur yang dimilikinya, salah satu fitur yang bisa dimanfaatkan untuk penelitian adalah ulasan online. Secara umum, proses menghasilkan setiap ulasan produk pada dasarnya terkait dengan tingkat rating(X. Wang et al., 2022). Misalnya, untuk pengguna yang toleran, meskipun pengguna sangat tidak puas dengan produk, namun perangkingan tetap membuatnya memberikan komentar netral yang tidak dapat menunjukkan kualitas produk. Ulasan dapat dimanfaatkan sebagai modal pokok utama dalam menentukan keputusan manajemen di perusahaan(Bi et al., 2019). Ulasan online sangat krusial ketika menentukan keputusan pembelian karena dengan berkembangnya e-commerce, semakin banyak ulasan palsu sehingga semakin banyak konsumen yang khawatir akan tertipu dalam belanja online(Q. Wang et al., 2022). Hal ini tidak bisa dipungkiri dikarenakan dari review pelanggan dapat ditentukan tingkat kepuasaan pengguna atau pembeli terhadap sebuah produk(Kevin et al., 2020).

Beberapa metode atau teknik untuk mengetahui respon terhadap suatu barang salah satunya adalah menggunakan analisis sentimen(Sistem et al., 2021). Seperti penelitian yang dilakukan oleh (Rohman et al., 2020) dan (Yennimar & Rizal, 2019) menggunakan metode Naïve Bayes dan KNN pada klasifikasi sentimen produk ecommerce. Algoritma Naïve Bayes mempunyai akurasi lebih tinggi dibanding dengan KNN. Selain menggunakan algoritma NB dan SVM juga mudah ditemui dalam mengambil keputusan terhadap suatu barang (Basani et al., 2019) dan klasifikasi ulasan produk di lima E-Commerce di Indonesia (Afifah et al., 2021). Ada teknik atau metode yang dapat dimanfaatkan untuk merepresentasikan teks menjadi bentuk vektor, salah satunya dengan menggunakan word embedding. Penelitian(Agustiningsih et al., 2022) menguji penggunaan word embedding pada klasifikasi sentimen menggunakan Model LSTM dua arah. Penelitian lain terkait word embedding yaitu(Nurdin et al., 2020) kinerja CNN dalam mengklasifikasikan teks ulasan menggunakan word embedding menggunakan ukuran *F Measure*.

Selain menggunakan *word embedding* untuk proses mengubah data teks mentah menjadi yang dapat dibaca mesin bisa menggunakan metode TF-IDF, penelitian(Afifah et al., 2021) menjelaskan bahwa kombinasi TF-IDF dengan Algoritma XGBoost menghasilkan akurasi 96,24%, studi ini membuktikan bahwa XGBoost cukup powerfull dalam menangani data tidak seimbang antar label. Hal ini sudah dibuktikan oleh penelitian(Akter et al., 2021) terkait Algoritma XGBoost juga membuktikan bahwa dari data pengolahan yang dilakukan oleh algoritma mesin dengan bobot kata TF-IDF dan algoritma klasifikasi XGBoost memiliki nilai F1-Score 0.91% lebih tinggi dibandingkan dengan TF-IDF Logistic Regression yang hanya memiliki 0.90%.

Berdasarkan penelitian sebelumnya, penelitian ini akan membandingkan kinerja Algoritma Naive Bayes, SVM dan XGBoost terhadap ulasan berbahasa Indonesia yang nantinya akan diuji dan dievaluasi hasilnya mana yang lebih baik dalam mengelompokkan data teks yang tidak seimbang antar labelnya.

2. Kajian Literatur

Karya Ilmiah (Rohman et al., 2020)melakukan penelitian terkait sentimen analisis terhadap produk dengan mengambil data sejumlah 3341 ulasan dengan hasil klasifikasi metode Naïve Bayes dengan Unigram akurasi 52,4% lebih rendah dibanding dengan

KNN dengan Unigram dengan nilai akurasi 76.2%. Untuk hasil akurasi kurang begitu optimal dikarenakan adanya ulasan yang tidak relevan terhadap suatu produk. Hal ini bisa dijadikan referensi bagi penulis saat ini yang akan meneliti topik yang berkaitan dengan ulasan produk dengan menambahkan proses cleaning data berupa menghapus setiap ada ulasan tidak relevan dan terdeteksi spam.

Hasil berbeda dikemukakan oleh penelitian(Yennimar & Rizal, 2019) dengan menggunakan algoritma sama bahwa NB menghasilkan performa akurasi 89.00% lebih baik dibanding KNN dengan nilai akurasi sebesar 67.00%. Perbedaan cukup signifikan terletak pada metode preprocessing yang digunakan dimana pada penelitian(Rohman et al., 2020) menggunakan penambahan proses normalizer kata. Normalizer kata sangat penting dilakukan di tahap preprocessing agar kata atau kosakata yang salah ketik atau tidak sesuai ejaan akan dibenarkan sehingga tidak hilang ketika proses di langkah lanjutan.

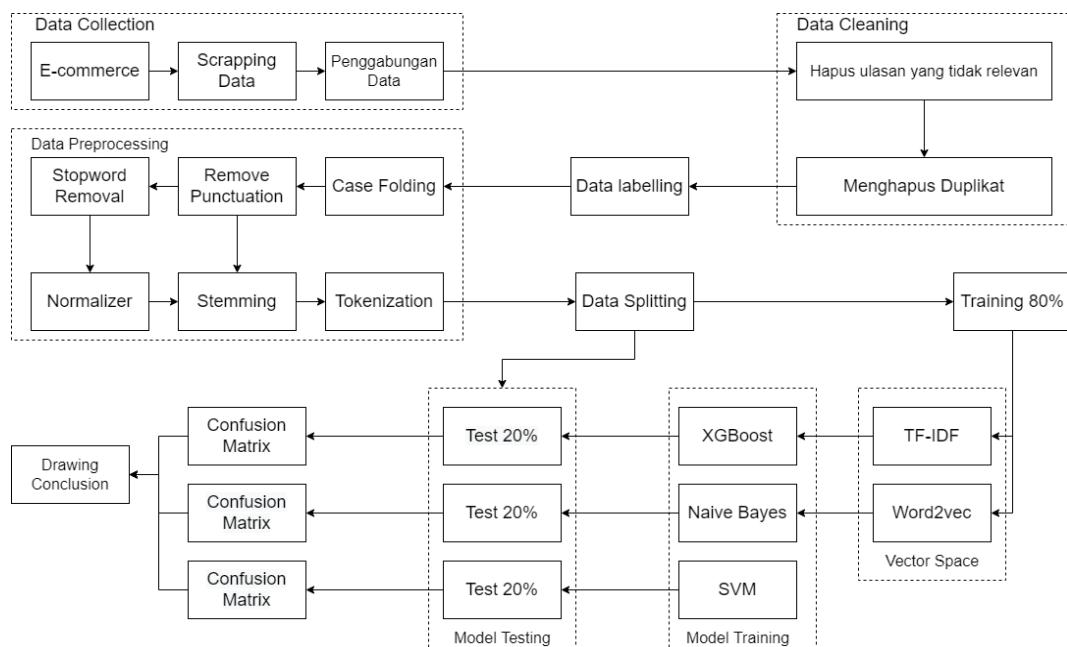
Preprocessing dalam sebuah analisis sentimen merupakan tahapan yang tidak bisa ditinggalkan pada proses mining, hal ini dikarenakan data yang digunakan pada proses penambangan data tidak semuanya dalam kondisi bisa diproses. Penelitian(Sihombing et al., 2021) melakukan klasifikasi pada ulasan online terhadap produk hp yang diperjual belikan pada web Shopee di Indonesia menggunakan algoritma Naïve Bayes, akurasi yang dihasilkan sebesar 85%, hasil ini belum bisa dikatakan optimal karena masih terdapat data misclassified. Hal ini bisa dijadikan acuan pada penelitian yang akan dilakukan perlu memperhatikan struktur dan metode pada tahap preprocessing agar kerusakan data seperti missing value, duplikasi, kesalahan system bisa diatasi.

Penelitian(Nurdin et al., 2020) membahas tentang kinerja algoritma CNN dalam mengklasifikasikan teks menggunakan word embedding. Hasil klasifikasi diukur menggunakan F-Measure secara berturut-turut adalah word2vec 0.925, glove 0.958, dan Fastext 0.979. Word2vec di penelitian sebelumnya tergolong mempunyai F-Measure yang lebih rendah dibanding Fasttext dan Glove. Hal ini membuat penulis terkait untuk mencoba memaksimalkan performa word2vec dengan kombinasi metode XGBoost dimana menurut(Afifah et al., 2021) algoritma ini cukup powerfull dalam mengatasi imbalance data. Hal ini diperkuat oleh percobaan yang telah dilakukan sebelumnya XGBoost menghasilkan akurasi 96,24% dengan dataset yang dikumpulkan dari Google Play Store.

Penelitian selanjutnya(Akter et al., 2021) membahas kinerja bobot kata TF-IDF pada kedua algoritma yaitu XGBoost dan algoritma klasifikasi Logistic Regression. Hasil uji coba pada penelitian ini mengemukakan bahwa dari data pembelajaran mesin dengan word2vec memiliki nilai F1-Score 0.91% lebih tinggi dibandingkan dengan TF-IDF Logistic Regression yang hanya memiliki 0.90%, hal ini menjadikan alasan utama penulis menggunakan algoritma ini karena XGBoost merupakan algoritma yang populer dan banyak digunakan karena memiliki karakteristik powerful. Berdasarkan penelitian sebelumnya, penelitian ini akan membandingkan kinerja Algoritma Naive Bayes, SVM dan XGBoost terhadap ulasan berbahasa Indonesia yang nantinya akan dievaluasi algoritma dengan kombinasi mana yang mampu memecahkan permasalahan imbalance data.

3. Metode Penelitian

Pada penelitian ini menggunakan 7 tahapan seperti yang di ilustrasikan di Gambar 1.



Gambar. 1 Metodologi Penelitian

Tahap pertama adalah Pengumpulan data dilakukan menggunakan metode scrapping di web Shopee Marketplaces , data yang digunakan hanya teks ulasan dan Rating dan disimpan dalam format .csv. Setelah itu tahap selanjutnya adalah pembagian data dimana data dibagi 80% data latih dan 20% data testing atau uji. Setelah data di split, data akan masuk proses cleaning data dimana ulasan-ulasan yang terindikasi spam dan tidak relevan akan dibersihkan. Setelah masuk tahap cleaning data, tahap selanjutnya preprocessing, tahap ini adalah tahap mengolah teks sebelum masuk di proses klasifikasi Untuk tahapan *preprocessing* bisa dilihat pada tabel 1.

Table 1.Tahap Preprocessing Teks Ulasan Produk

No	Tahap Preprocessing	Proses
1	Teks Kotor	Produk ini jlek bangett ☺:(aku kecewa sama yang ini udah membeli disini ni...
2	Case Folding	produk ini jlek bangett ☺:(aku kecewa sama yang ini udah membeli disini ni...
3	Remove Punctuation	produk ini jlek bangett aku kecewa sama yang ini udah membeli disini ni
4	Remove Number & short word	produk jlek bangett kecewa sama yang udah membeli disini
5	Word normalize	produk jelek banget kecewa sama yang udah beli disini

6	Stopword removal	produk jelek banget kecewa sama udah membeli disini
7	Tokenization	“produk”, “jelek”, “banget”, “kecewa”, “sama”, “udah”, “membeli”, “disini”
8	Stemming	“produk”, “jelek”, “banget”, “kecewa”, “sama”, “udah”, “beli”, “sini”

Setelah ulasan selesai dipreprocessing tahap selanjutnya adalah pelatihan model. Ada 6 Skema model pelatihan yaitu TF-IDF+Naïve Bayes, Word2vec+NaiveBayes, TF-IDF+SVM, Word2vec+SVM, TF-IDF+XGBoost dan Word2vec+XGBoost. Tahap terakhir adalah evaluasi dimana hasil klasifikasi dari enam model tersebut akan dievaluasi menggunakan confusion matrix. Menurut(Kotu & Deshpande, 2019) confusion matrix merupakan sebuah tabulasi dari perhitungan untuk menentukan benar atau salah pada suatu class..

4. Hasil dan Pembahasan

Proses scrapping data pada web shopee dibagian ulasan produk menghasilkan menghasilkan 25.581 data yang terbagi 80% data latih dan 20% data uji. Dalam data latih diketahui label baik memiliki jumlah 8141 sedangkan untuk label buruk ada 2323, ini adalah tantangan klasifikasi yang tidak seimbang. Tahap selanjutnya data teks akan memasuki tahap cleaning data dimana data yang spam dan tidak relevan akan dihilangkan, jadi data yang mengandung kata kunci produk local saja yang akan diproses, setelah cleaning data selesai tahap berikutnya adalah labeling dimana tahap ini jika suatu ulasan mempunyai rating 1 – 3 maka akan di labelkan negative atau 0 dan jika ulasan memiliki rating 4-5 maka akan diberikan label positif atau 1. Setelah selesai proses labeling tahap selanjutnya adalah proses preprocessing.. Berikut adalah teks ulasan yang sudah melewati proses preprocessing akan disajikan dalam bentuk worcloud pada gambar 2.



Gambar 2. Worcloud Ulasan Produk

Tahap selanjutnya adalah tahap evaluasi sebuah model terhadap metode yang digunakan. Ke Enam model yang sudah dilatih menggunakan data latih akan diukur menggunakan F1-Score dapat dilihat di tabel 2.

Table 2.Evaluasi Model menggunakan F1-Score

	Naïve Bayes	SVM	XGBoost
TF-IDF	0.915	0.938	0.940
Word2vec	0.900	0.939	0.941

Tabel 2 menjelaskan bahwa kombinasi word embedding Word2vec + algoritma klasifikasi XGBoost menghasilkan F1 Score lebih bagus yaitu sebesar 0.941 diikuti dengan TF-IDF + XGBoost 0.940. Sementara untuk algoritma SVM dengan menggunakan vector space TF-IDF dan Word2vec hanya menghasilkan 0.938 dan 0.939. Sedangkan untuk Naïve Bayes memiliki F1-Score 0.915 dengan TF-IDF dan 0.900 dengan word2vec. Klasifikasi dengan word2vec dalam merepresentasikan kata menjadi vektor lebih baik dibandingkan dengan TF-IDF, hal ini dikarenakan word2vec mampu memproses hubungan dekat antar kata. Dapat dilihat di tabel 3 bagaimana word2vec memproses relasi semantik antar kata.

Table 3.Relasi semantic antar kata pada word2vec

Kata “Kecil”	Nilai Kedekatan
Sempit	0.591
Ngepas	0.582
Cingkrang	0.577
Badan	0575
Ngetat	0575

Pada tabel 3. Bisa diketahui bahwa kedekatan kata kecil mempunyai relasi semantik dengan kata sempit dengan score 0.591 diikuti dengan kata ngepas,cingkrang,badan dan ngetat. Hal lain yang bisa didapatkan dari penelitian ini adalah bahwa model TF-IDF+XGBoost pada penelitian yang dilakukan memiliki performa uji nilai F1-Score lebih baik 0,940% dibandingkan penelitian yang dilakukan sebelumnya dengan model yang sama yang dilakukan oleh (Akter et al., 2021) TF-IDF + XGBoost memiliki nilai F1-Score 0.91%. Hal ini membuktikan bahwa metode penambahan *word normalizer* atau normalisasi kata terbukti bisa menaikkan nilai F1 Score dalam model algoritma klasifikasi.

5. Kesimpulan

Analisis sentimen dapat diterapkan untuk mengetahui kepuasan pengguna atau pembeli terhadap suatu produk, rekomendasi yang dihasilkan dari analisis sentimen dapat menjadi perbaikan dan saran untuk penjual agar bisa memberikan kualitas produk yang baik. Berdasarkan penelitian yang telah dilakukan, bahwa kombinasi Word2vec + XGBoost menghasilkan F1 Score lebih tinggi yaitu sebesar 0.941 diikuti dengan TF-IDF + XGBoost 0.940. Sementara untuk algoritma SVM dengan menggunakan vector space TF-IDF dan Word2vec hanya menghasilkan 0.938 dan 0.939. Sedangkan untuk Naïve

Bayes memiliki F1-Score 0.915 dengan TF-IDF dan 0.900 dengan word2vec. Klasifikasi dengan word2vec dalam merepresentasikan kata menjadi vektor lebih baik dibandingkan dengan TF-IDF, hal ini dikarenakan kelebihan word2vec mampu mengolah kata lebih baik berdasarkan kategori semantic atau kedekatan kata.

Dalam penelitian selanjutnya, akan lebih baik untuk menambahkan deteksi sarkasme dan pelabelan otomatis agar tidak adanya bias dan mengklasifikasikan ulasan berdasarkan topic tidak hanya menggunakan sentimen.

Daftar Pustaka

- Afifah, K., Yulita, I. N., & Sarathan, I. (2021). Sentiment Analysis on Telemedicine App Reviews using XGBoost Classifier. *2021 International Conference on Artificial Intelligence and Big Data Analytics*, 22–27. <https://doi.org/10.1109/ICAIBDA53487.2021.9689735>
- Agustiningsih, K. K., Utami, E., Muhammad, O., & Alsyabani, A. (2022). *Sentiment Analysis of COVID-19 Vaccines in Indonesia on Twitter Using Pre-Trained and Self-Training Word Embeddings*. 1, 39–46.
- Akter, M. T., Begum, M., & Mustafa, R. (2021). Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors. *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 40–44. <https://doi.org/10.1109/ICICT4SD50815.2021.9396910>
- Basani, Y., Sibuea, H. V., Sianipar, S. I. P., & Samosir, J. P. (2019). Application of Sentiment Analysis on Product Review E-Commerce. *Journal of Physics: Conference Series*, 1175, 12103. <https://doi.org/10.1088/1742-6596/1175/1/012103>
- Bi, J.-W., Liu, Y., & Fan, Z.-P. (2019). Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking. *Information Sciences*, 504, 293–307. <https://doi.org/https://doi.org/10.1016/j.ins.2019.07.025>
- Kevin, V., Que, S., Iriani, A., & Purnomo, H. D. (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), 162–170.
- Kotu, V., & Deshpande, B. (2019). Chapter 8 - Model Evaluation. In V. Kotu & B. Deshpande (Ed.), *Data Science (Second Edition)* (Second Edition, hal. 263–279). Morgan Kaufmann. <https://doi.org/https://doi.org/10.1016/B978-0-12-814761-0.00008-3>
- Nurdin, A., Seno aji, B., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, 14, 74. <https://doi.org/10.33365/jtk.v14i2.732>
- Rohman, A. N., Luviana Musyarofah, R., Utami, E., & Raharjo, S. (2020). Natural Language Processing on Marketplace Product Review Sentiment Analysis. 2020

2nd International Conference on Cybernetics and Intelligent System (ICORIS), 1–5.
<https://doi.org/10.1109/ICORIS50180.2020.9320827>

Sihombing, L., Hannie, H., & Dermawan, B. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier. 5, 233–242. <https://doi.org/10.29408/edumatic.v5i2.4089>

Sistem, R., Lestandy, M., Abdurrahim, A., & Syafa, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent. 5(10), 802–808.

Wang, Q., Zhang, W., Li, J., Mai, F., & Ma, Z. (2022). Effect of online review sentiment on product sales: The moderating role of review credibility perception. *Computers in Human Behavior*, 133, 107272. <https://doi.org/https://doi.org/10.1016/j.chb.2022.107272>

Wang, X., Zhou, T., Wang, X., & Fang, Y. (2022). Harshness-aware sentiment mining framework for product review. *Expert Systems with Applications*, 187, 115887. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115887>

Yennimar, Y., & Rizal, R. (2019). Comparison of Machine Learning Classification Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency. *SinkrOn*, 4, 268. <https://doi.org/10.33395/sinkron.v4i1.10416>

Simon Kemp, 2022, Digital 2022 Indonesia :Internet use in Indonesia 2022, <https://datareportal.com/reports/digital-2022-indonesia?rq=indonesia%202022>, diakses tanggal 5 April 2022

Kominfo.Kulonprogo, 2022, Presiden: Bangga Gunakan Produk Indonesia,<https://kominfo.kulonprogokab.go.id/detil/1191/presiden-bangga-gunakan-produk-indonesia>, diakses tanggal 5 April 2022

Ekonomi.Bisnis.com, 2022, Simak! Ini 3 Alasan Beli Produk Lokal Selama Ramadan, <https://ekonomi.bisnis.com/read/20220405/12/1519120/simak-ini-3-alasan-beli-produk-lokal-selama-ramadan>, diakses tanggal 5 April 2022