

Analisis Sentimen *E-Commerce* Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma *Support Vector Machine*

Anggi Jiana Putri¹, Ardia Salsabila Syafira², Muhammad Eka Purbaya^{*3}, Denny Purnomo⁴

^{1,2} Program Studi Sistem Informasi ITTP, ^{*3,4} Program Studi Bisnis Digital ITTP

Jl. DI. Pandjaitan No.128 Purwokerto, Banyumas, 53147, Jawa Tengah, Indonesia

Email: 18103036@ittelkom-pwt.ac.id¹, 18103037@ittelkom-pwt.ac.id², m.eka@ittelkom-pwt.ac.id³, denny@ittelkom-pwt.ac.id⁴

Received: Jan 23, 2022 / Revised: Feb 25, 2022 / Accepted: March 11, 2022

Abstrak

E-commerce dapat mempermudah transaksi jual beli antara penjual dan pembeli melalui sebuah platform online. Perkembangan teknologi yang semakin cepat secara tidak langsung memaksa pengusaha untuk mengembangkan sayapnya melalui *e-commerce*. Lazada merupakan perusahaan *e-commerce* terbesar no. 4 yang ada di Indonesia. Bagi perusahaan sangat penting untuk mengetahui penilaian mengenai layanan yang mereka tawarkan. Salah satu media yang biasa digunakan untuk mengutarakan opini kita adalah melalui media sosial yang salah satunya adalah *Twitter*. Pendapat dan komentar tentang *e-commerce* Lazada di *Twitter* sangat beragam. Dalam rangka meningkatkan pelayanannya perusahaan maka perlu mengkaji pendapat publik terhadap *e-commerce* Lazada. Pendekatan yang dapat dilakukan untuk mengkaji pendapat publik adalah melalui analisis sentimen. Algoritma *Support Vector Machine* (SVM) diketahui dapat melakukan analisis sentimen dengan baik. Berdasarkan penelitian yang telah dilakukan, metode *vader* yang ada dalam *library* Orange Data Mining dapat membantu pelabelan dataset 300 *tweet*. Performa algoritma *Support Vector Machine* dalam proses *training dan testing* telah menghasilkan nilai *accuracy*, *precision*, *recall* dan *F1 score* diatas 80 %. Hasil analisis sentimen dari penelitian ini dapat digunakan sebagai tolak ukur untuk peningkatan layanan perusahaan dan memberikan wawasan kepada masyarakat.

Kata kunci: analisis sentimen, *e-commerce*, lazada, twitter, *support vector machine*.

Abstract

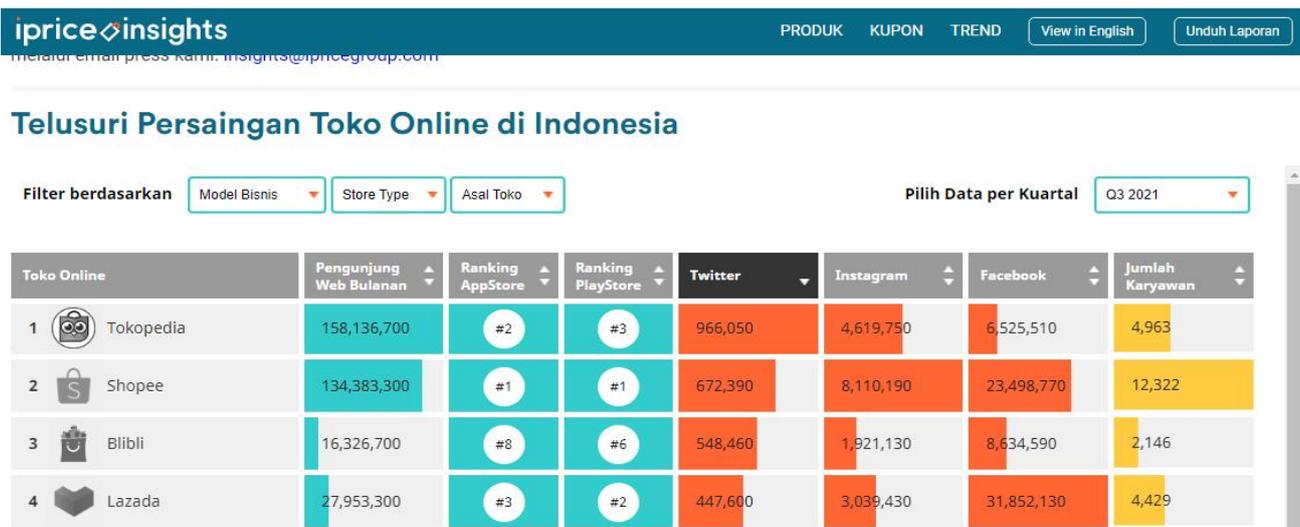
E-commerce can facilitate buying and selling transactions between sellers and buyers through an online platform. Rapid technological developments indirectly force entrepreneurs to develop their wings through *e-commerce*. Lazada is the largest *e-commerce* company no. 4 in Indonesia. For companies it is very important to know the assessment of the services they offer. One of the media commonly used to express our opinions is through social media, one of which is *Twitter*. Opinions and comments about Lazada *e-commerce* on *Twitter* vary widely. In order to improve the company's services, it is necessary to examine public opinion on Lazada's *e-commerce*. The approach that can be taken to examine public opinion is through sentiment analysis. The *Support Vector Machine* algorithm is known to be able to perform sentiment analysis well. Based on the research that has been done, the *Vader* method in the *Orange Data Mining library* can help label the 300 *tweet* dataset. The performance of the *Support Vector Machine* algorithm in the *training and testing* process has resulted in an *accuracy* value of 86.7%; *precision* 86,2 % ; *recall* 86.7% and *F1 score* 84.9%. The results of sentiment analysis from this study can be used as a benchmark for improving company services and providing insight to the public.

Keywords: *e-commerce*, lazada, sentiment analysis, *support vector machine*, twitter.

1. Pendahuluan

Perkembangan teknologi yang semakin cepat secara tidak langsung memaksa pengusaha-pengusaha yang bergerak di bidang ritel untuk mengembangkan sayapnya mengarahkan pemasaran maupun penjualan melalui *e-commerce*. *E-commerce* dapat mempermudah transaksi jual beli antara penjual dan pembeli melalui sebuah *platform* online. *E-commerce* terdapat dalam bentuk website dan juga aplikasi mobile sehingga

memungkinkan masyarakat atau pengguna dapat dengan mudah mengaksesnya di manapun dan kapanpun. *E-commerce* Lazada adalah salah satu *e-commerce* yang marak diiklankan di media sosial salah satunya di *Twitter*. Menurut SimilarWeb pada laman iprice.co.id (Iprice, 2022), seperti yang terlihat pada gambar 1, Lazada masuk pada peringkat ke-4 dalam cuitan yang ada dalam *Twitter* pada kuartal 3 tahun 2021.



Gambar 1. Perbandingan E-Commerce menurut <https://iprice.co.id/insights/mapofecommerce/>

Lazada merupakan perusahaan *e-commerce* swasta yang berasal dari Singapura yang didirikan oleh Rocket Internet yang menawarkan berbagai jenis produk mulai dari elektronik, fashion wanita & pria, peralatan rumah tangga, kesehatan & kecantikan, mainan bayi & anak, olahraga & travel, grosir, otomotif & media. Selain itu juga, Lazada juga memiliki banyak metode pembayaran yang membuat pelanggan dapat leluasa memilih metode pembayaran yang mereka inginkan seperti bank transfer, dana, ovo, bayar di counter (Alfamart & Indomaret), *cash on delivery*, kartu kredit & debit, lazada credit, dan transaksi cicilan menggunakan aplikasi yang tersedia (kredivo dan akulaku).

Bagi perusahaan atau organisasi sangat penting untuk mengetahui penilaian mengenai layanan yang mereka tawarkan. Pelanggan biasanya menggunakan media sosial untuk melampiaskan perasaannya terhadap layanan yang diberikan oleh perusahaan *e-commerce*. Banyak media sosial yang pelanggan gunakan untuk mengutarakan pendapatnya. Salah satu media yang biasa digunakan yaitu Twitter. Twitter merupakan situs website layanan jejaring sosial sebagai media komunikasi dan mendapatkan informasi berupa pertanyaan, pendapat atau komentar. Pendapat dan komentar tentang *e-commerce* Lazada di Twitter sangat beragam. Dalam rangka meningkatkan pelayanannya perusahaan maka perlu mengkaji pendapat publik terhadap *e-commerce* Lazada. Oleh karena itu, untuk mengkaji pendapat publik terhadap *e-commerce* Lazada, pada penelitian ini menggunakan analisis sentimen mengambil data *tweet* masyarakat yang berkomentar tentang *e-commerce* Lazada.

Penelitian ini menggunakan beberapa referensi penelitian sebelumnya mengenai sentimen analisis. Pravina et al., (2019) melakukan analisis sentimen mengenai opini maskapai penerbangan menggunakan

metode SVM. Dengan menggunakan tahap *pre-processing* seperti *case-folding*, *cleaning*, *translation*, *stemming*, *stopword removal* dan tokenisasi yang telah dilakukan Pravina et al., (2019) memberikan hasil *accuracy* sebesar 40%, *precision* 40%, 100% *recall*, dan *f-measure* sebesar 57,14%.

Sedangkan pada penelitian yang dilakukan oleh Pamungkas et al., (2021) yang meneliti data *tweet* dengan topik Kasus Benih Lobster 2020 menghasilkan nilai akurasi akurasi 84,21% sedangkan nilai sensitivitas 73,38% dan nilai spesifisitas 82,10% menggunakan metode SVM untuk sentimen analisis. Dan selanjutnya adalah penelitian dari Luqyana et al., (2018) yang melakukan analisis sentimen Cyberbullying pada komentar instagram menggunakan metode *Support Vector Machine*. Hasil yang diperoleh pada penelitian tersebut adalah metode SVM menghasilkan akurasi sebesar 90%, *precision* sebesar 94,44%, 85% *recall* sebesar dan *f-measure* sebesar 89,47%.

Berdasarkan studi literatur pada penelitian sebelumnya, maka pada penelitian ini akan menggunakan algoritma *Support Vector Machine* dengan tujuan untuk mengklasifikasi atau mengelompokkan sentimen positif dan negatif pengguna terhadap layanan *e-commerce* Lazada. Hasil dari penelitian ini dapat diketahui persentase sentimen publik terhadap *e-commerce* Lazada yang telah diklasifikasikan ke dalam kelompok positif dan negatif sehingga dapat diketahui reaksi publik terhadap layanan yang telah diberikan dan dapat digunakan sebagai tolak ukur untuk peningkatan layanan perusahaan dan memberikan wawasan kepada masyarakat.

2. Metode Penelitian

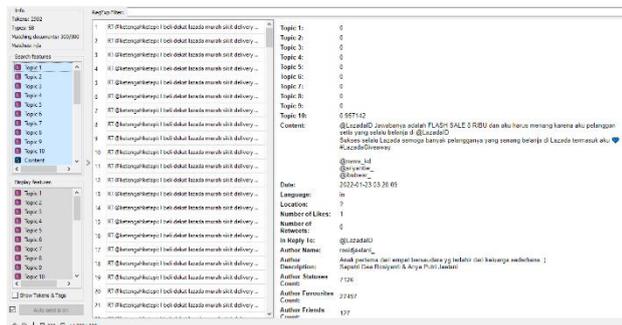
Dalam penelitian ini terdapat beberapa tahap, antara lain pengumpulan data mentah yang diperoleh dari proses *crawling* data jejaring sosial *twitter*

3. Hasil dan Pembahasan

Berikut merupakan hasil dan pembahasan dari penelitian ini:

3.1 Pengumpulan Data

Sumber data teks yang akan diolah pada penelitian ini berasal dari data *tweet* maupun *retweet* yang diambil langsung dari jejaring sosial memanfaatkan API key yang telah disediakan oleh *twitter*. Jumlah data yang diambil pada penelitian ini berjumlah 300 data *tweet* yang dilakukan pada tanggal 21 Januari 2022 dengan hasil *crawling* data seperti yang dilihat pada gambar 3.

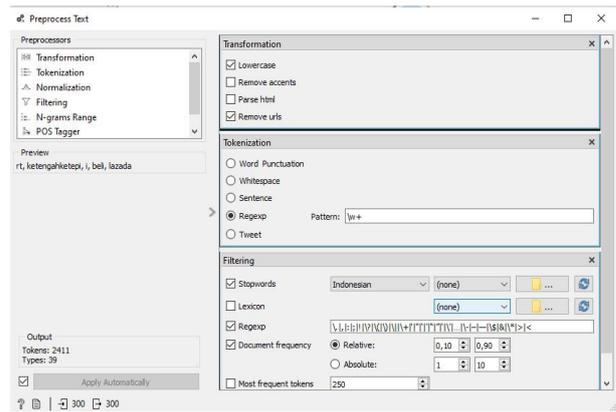


Gambar 2. Hasil pengumpulan data

3.2 Pre-processing

Pre-processing dilakukan dengan metode yang sudah ditentukan sebelumnya yaitu: *cleansing*, *case folding*, *tokenizing*, *normalisasi*, *filtering* dan *stemming*. *Cleansing* dilakukan untuk membersihkan karakter yang tidak digunakan seperti URL, @, #, dan simbol yang tidak digunakan pada proses selanjutnya. Setelah menghilangkan karakter yang tidak terpakai, dilakukan *case folding* untuk membuat semua karakter menjadi huruf kecil.

Kemudian dilakukan *tokenizing* yang memecah teks menjadi komponen-komponen yang lebih kecil yaitu kata, kalimat akan membagi teks kata demi kata dan meninggalkan simbol tanda baca juga. Kemudian *normalisasi*, teks yang telah dipecah akan menghasilkan teks yang semula tidak baku menjadi baku. Langkah selanjutnya adalah *filtering* untuk menghilangkan tanda baca yang tentunya tidak memiliki arti dan mengaktifkan *document frequency* agar lebih tersaring lagi kata-kata yang terlalu sering muncul dan terlalu jarang muncul agar tidak dimasukkan dalam kamus kata kita. Pengaturan *pre-processing* ini dapat dilihat pada gambar 3.



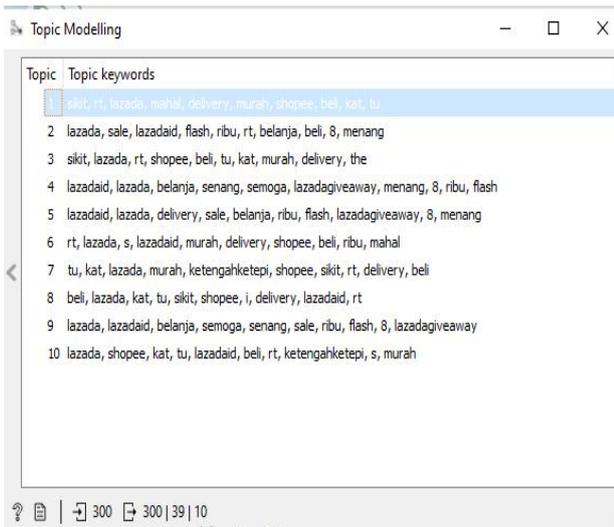
Gambar 3. Proses pre-processing text

Jika tahap *pre-processing* telah dilakukan, akan terbentuk data teks-teks terpisah dan dapat dilihat dalam bentuk *word cloud* pada Orange. Fungsi *widget word cloud* membagikan tampilan dengan frekuensi jumlah kata yang paling banyak. Jika semakin banyak frekuensi kemunculan kata, ukuran huruf pada *word cloud* menjadi besar. Gambar 4 merupakan hasil dari tampilan *word cloud*, dapat diambil kesimpulan bahwa topik yang berkaitan dengan Lazada adalah sikit, lazadaid, shopee.



Gambar 4. Tampilan hasil word cloud

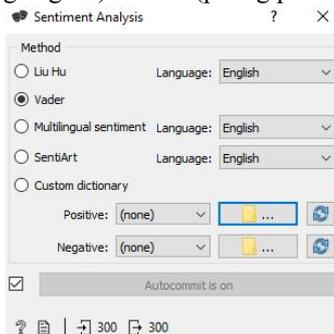
Untuk lebih memperkuat analisis topik apa saja yang berkaitan dengan Lazada, dapat digunakan *widget topic modelling* dengan mengimplementasi metode *Latent Dirichlet Allocation (LDA)*. Hasil dari *topic modelling* dapat dilihat pada gambar 5.



Gambar 5. Hasil *topic modelling*

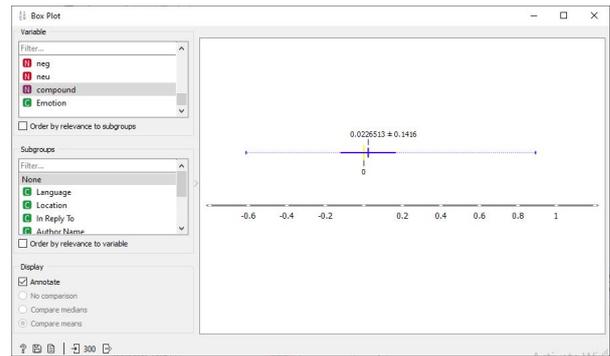
3.3 Analisis Sentimen

Pada tahap ini dilakukan analisis sentimen untuk mengelompokkan masing-masing *tweet* pada agar dapat langsung dikategorikan sebagai opini positif, negatif atau netral dengan menggunakan metode *Vader* yang sudah disediakan oleh *Orange Data Mining*. *VADER* (*Valence Aware Dictionary dan Sentiment Reasoner*) merupakan *library* penganalisis sentimen sumber terbuka berbasis leksikon. Dengan mengimplementasikan *vader* seperti yang ditunjuk gambar 6, hasil akhir yang diperoleh adalah akan ada penambahan kolom positif, negatif, netral dan *compound* (polaritas keseluruhan kalimat). *Compound* merupakan metrik yang menghitung jumlah semua peringkat *lexicon* yang telah dinormalisasi dengan antara -1 (paling negatif) dan +1 (paling positif).



Gambar 6. Pengaturan analisis sentimen

Dengan memanfaatkan *widget box plot*, kita bisa melihat pemetaan yang dihasilkan dalam sentimen analisis dari 300 data *twitter*. Pada gambar 7 telah disajikan pemetaan rerata hasil opini melalui *field compound* dengan nilai $0,023 \pm 0,14$ yang mengartikan bahwa secara umum opini masyarakat terhadap *e-commerce* Lazada bersifat positif.



Gambar 7. Grafik *compound*

Vader digunakan untuk membantu memberikan label pada setiap *tweet* pada set data. Data terbaru kemudian disimpan dengan tipe file *excel* yang nantinya akan dilakukan transformasi nilai dari kolom *compound* menjadi label dari masing-masing *tweet* yakni positif, negatif dan netral seperti yang disajikan pada gambar 8.

3.4 Klasifikasi *Support Vector Machine*

Data sentimen analisis dengan ekstensi *xls* yang telah dilakukan pelabelan, lalu di-*import* melalui *widget corpus* di *Orange Data Mining*. Selanjutnya digunakan *select column* untuk memilih kolom Sentimen dan Content saja yang akan digunakan untuk analisis selanjutnya. Data tabel digunakan untuk menampilkan data *tweet* berdasarkan kolom yang telah kita pilih seperti yang terlihat pada gambar 8.

Sentimen	Content
Positif	RT @jeonghaniyou: update: alhamdulillah sisa 4 lagi yuk bisa! 🙏🙏🙏🙏🙏🙏
Positif	@armyfes: Lazada always lazadaa :)
Positif	Ada yg mau sharing sama aku ga ya? Cuma mau photocard sama albumnya aja. Pengiriman pak...
Positif	@LazadaID LAZADA TIPU! DUIT PENCAIRAN KK GUA 13 JUTA GA DIBAYAR2 UDAH 1 BULANANI ...
Positif	Nak shopping Shopee, Lazada, FoodPanda, Grab jangan lupa pakai shopback !...
Positif	RT @jeonghaniyou: update: alhamdulillah sisa 4 lagi yuk bisa! 🙏🙏🙏🙏🙏🙏
Positif	RT @LazadaID: Penampilan kamu nggak akan lengkap tanpa ditemani produk ternama dari Nike ...
Positif	shopee kedekut free shipping sekarang...
Positif	#realme GT Master Edition 8128GB 5G By Lazada Superiphone 242...
Positif	1. Pengenalan Lazada ...
Positif	https://t.co/CJqa5jOQta ==> Pemerah Bibir Alami Original Beauty Clinic #pemerahbibir ...
Positif	RT @jeonghaniyou: update: alhamdulillah sisa 4 lagi yuk bisa! 🙏🙏🙏🙏🙏🙏

Gambar 8. Data Tabel

Lalu data sentimen tersebut dilakukan *pre-processing* lagi sama seperti sebelumnya, yang membedakan disini adalah perlu ditambahkan komponen filter *most frequent tokens* agar bisa dibatasi kata apa saja yang penting untuk diolah. Selanjutnya dilakukan *bag of words* untuk mengubah data teks menjadi vektor yang dapat dipahami oleh komputer. Metode ini membantu menghitung frekuensi kemunculan kata pada seluruh dokumen.

Setiap frekuensi kemunculan kata dihitung dan diberikan kemungkinan probabilitas kemunculan kelas positif, negatif dan netral. Untuk melakukan prediksi bisa dilakukan dengan mencocokkan setiap kata pada data testing lalu dihitung probabilitas per kata dalam *corpus* tersebut. Ketiga kelas tersebut akan dibandingkan mana yang tertinggi untuk memberikan

keputusan bahwa sebuah *tweet* bisa diklasifikasikan masuk ke kelas positif, negatif atau netral.

Proses selanjutnya adalah tahapan *training*, *testing* menggunakan model *Support Vector Machine*. Teknik *sampling* yang digunakan untuk pelatihan adalah menggunakan metode *Cross Validation* dengan nilai *k-fold* 10. Melalui bantuan *widget confusion matrix*, dapat dilakukan evaluasi performa algoritma *Support Vector Machine* menggunakan acuan *confusion matrix*. Berdasarkan hasil *confusion matrix* seperti yang disajikan pada gambar 9, kita dapat menghitung dan menentukan *accuracy*, *precission* dan *recall*. *Widget test & F1 score* secara otomatis sudah menghitung *accuracy*, *precision*, *recall* dan *F1 score* dengan hasil sebagai berikut :

- *Accuracy* : 86,7 %
- *Precision* : 86,2 %
- *Recall* : 86,7 %
- *F1 score* : 84,9 %

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	0	8	0	8
	Netral	0	198	0	198
	Positif	0	32	62	94
Σ		0	238	62	300

Gambar 9. Hasil *Confusion Matrix*

4. Kesimpulan

Berdasarkan hasil uraian yang telah dibahas, kesimpulan yang dapat ditarik pada penelitian ini adalah sentimen analisis dapat dilakukan menggunakan metode *vader* untuk membantu memberikan label pada data *tweet* yang telah di *crawling*. Model SVM juga bisa digunakan untuk tugas sentimen analisis. Berdasarkan hasil *accuracy*, *precission* dan *recall*. *widget test & F1 score* yang didapat menggunakan *tools Orange Data Mining* mampu menghasilkan nilai diatas 80 % sehingga dapat dikatakan algoritma *support vector machine* mampu melakukan analisis sentimen pada *e-commerce* Lazada.

Rekomendasi untuk penelitian selanjutnya, dataset yang sudah diolah dapat diproses dengan pemodelan lain maupun algoritma *TF-IDF (Term Frequency – Inverse Document Frequency)* untuk menganalisis hubungan antara sebuah frase/kalimat dengan sekumpulan dokumen karena pada penelitian ini hanya menggunakan mode *default* dari aplikasi *Orange Data Mining*.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Pengelola Prodi Sistem Informasi ITTP yang telah memfasilitasi penelitian ini melalui mata kuliah

Analitika Media sehingga tulisan dari luaran penelitian ini bisa dipublikasikan, serta Dosen Pengampu matakuliah Analitika Media dalam membantu kelancaran penelitian ini.

Daftar Pustaka

- Iprice. (2022). *Peta E-Commerce Indonesia*. <https://iprice.co.id/insights/mapofecommerce/>
- Luqyana, W. A., Gholissodin, I., & Perdana, rizal setya. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(11), 4704–4713. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto). *Jurnal Informatika dan Rekayasa Elektronik*, 2(2), 43. <https://doi.org/10.36595/jire.v2i2.117>
- Nurhafida, S. I., & Sembiring, F. (2021). Analisis Text Clustering Masyarakat di Twiter Mengenai Mcdonald'sxsbts Menggunakan Orange Data Mining. *Seminar Nasional Sistem Informasi dan Manajemen Informatika*, 1(1), 28–35.
- Pamungkas, B., Purbaya, M. E., & A.K, D. J. (2021). Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 3(2), 10–20. <https://doi.org/10.20895/inista.v3i2.243>
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2789–2797. <http://j-ptiik.ub.ac.id>
- Widagdo, A. S., W.A, B. S., & Nasiri, A. (2020). Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes. *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, 6(1), 1–5. <https://doi.org/10.46808/informa.v6i1.159>
- Yasid, M. (2019). Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes. *Jurnal Ilmiah Informatika*, 7(02), 82. <https://doi.org/10.33884/jif.v7i02.1329>