Universitas Muhammadiyah Sidoarjo

Seminar Proposal

Analisis Sentimen Publik Di Media Sosial TikTok Terhadap Penghentian Layanan J&T Express Pada Shopee Menggunakan Metode Suport Vector Machine (SVM)

Disusun Oleh :

Fathur Rahman Maulana (226120900001)

Dosen Pembimbing

Istian Kriya Almanfaluti,S.Kom., M.Kom

Dosen Penguji 1

Bayu Hari Prasojo, S.Si., M.Pd

Dosen Penguji 2

Andry Rachmadany, S.Kom., M.Kom

Program Studi Bisnis Digital Fakultas Bisnis Hukum Dan Ilmu Sosial Universitas Muhammadiyah Sidoarjo

2025-2026

Latar Belakang

Perkembangan teknologi digital mendorong pertumbuhan e-commerce yang sangat pesat di Indonesia. Salah satu platform yang memiliki dominasi cukup tinggi adalah Shopee. Dalam operasionalnya, e-commerce sangat bergantung pada sistem logistik untuk menjaga kepuasan pengguna.

Belakangan ini muncul isu penghentian kerja sama antara Shopee dan J&T Express. Isu tersebut menimbulkan berbagai reaksi dari masyarakat, terutama di media sosial seperti TikTok. Persepsi publik terhadap kebijakan ini berpotensi mempengaruhi reputasi perusahaan.

Berdasarkan studi literatur, penelitian analisis sentimen sebelumnya banyak menggunakan data dari Twitter atau Google Play Store. Namun, analisis berbasis komentar TikTok masih relatif jarang dilakukan. Selain itu, belum ditemukan penelitian yang secara spesifik membahas isu pemutusan mitra logistik ini menggunakan pendekatan SVM

Rumusan Masalah

Dampak dan pengaruh isu pemutusan hubungan kerja sama antara Shopee dan salah satu mitra logistik utamanya J&T Express

Pertanyaan :

- Bagaimana tanggapan masyarakat di media sosial TikTok terhadap isu pemutusan kerja sama antara Shopee dan J&T Express?
- Bagaimana proses pengelompokan sentimen masyarakat (positif, negatif, atau netral) terhadap isu tersebut dengan menggunakan metode Support Vector Machine?
- Apa dampak dari hasil analisis sentimen publik tersebut terhadap citra dan reputasi Shopee sebagai platform e-commerce?

Kategori SDGs

Berdasarkan Pendahuluan dan rumusan masalah yang dibahas dalam penelitian ini, maka fokus utamanya adalah pada kemitraan untuk mencapai tujuan yang tercantum pada nomor 17 yang mengarah pada transfer teknologi, pengembangan kapasitas, serta dukungan keuangan untuk memastikan keberlanjutan.

Literatur Review

- **E-commerce** merupakan platform digital yang memfasilitasi aktivitas transaksi jual beli secara online antara penjual dan pembeli. Dalam operasionalnya, e-commerce tidak hanya bergantung pada sistem aplikasi, tetapi juga sangat bergantung pada sistem logistik sebagai pendukung distribusi barang kepada konsumen.
- **Mitra Logistik.** Logistik digital meliputi proses penerimaan, penyortiran, pengiriman, hingga pelacakan paket, yang di mana mitra logistik menjadi faktor penentu pengalaman pengguna. Sehingga apabila terjadi keterlambatan pengiriman atau sistem tracking yang tidak akurat dapat langsung menurunkan kualitas layanan.
- **Sentimen Publik**. Bagian dari pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi suasana hati atau opini seseorang terhadap suatu topik melalui teks. Pada proses ini, sistem akan mengumpulkan dan menganalisis berbagai bentuk pendapat mulai dari ulasan, komentar, hingga unggahan media sosial untuk kemudian dikelompokkan ke dalam sentimen positif, negatif, atau netral.

Metode

Penelitian ini memakai metode Support Vector Machine (SVM) sebagai teknik utama dalam proses klasifikasinya. Support Vector Machine (SVM) merupakan salah satu metode supervised learning yang banyak dipakai dalam klasifikasi teks karena performanya yang cukup stabil dan akurasinya yang umumnya tinggi. Secara umum, ada tujuh tahapan yang dilakukan.



Metode

- Crawling Data

Sumber data penelitian ini diperoleh melalui proses crawling pada platform TikTok, khususnya pada dua unggahan video yang membahas isu pemutusan kerja sama antara Shopee dan J&T Express. Dari proses tersebut, diperoleh sebanyak 5.760 komentar.

Pembagian dataset pada penelitian ini dilakukan menggunakan pendekatan holdout method, yaitu dengan memisahkan data ke dalam training set dan testing set. Proporsi 80% data latih dan 20% data uji dipilih untuk memperoleh keseimbangan antara jumlah data yang memadai dalam proses pembelajaran model dan keandalan evaluasi performa.

- Pra-Pemrosesan Data

Pada tahap preprocessing, data komentar yang diperoleh terlebih dahulu dibersihkan melalui proses cleansing untuk menghapus elemen yang dianggap sebagai noise, seperti username, karakter HTML, emoji, hashtag, URL, dan simbol yang tidak relevan. Selanjutnya dilakukan case folding dengan menyeragamkan teks menjadi huruf kecil. Data kemudian diproses melalui tokenizing untuk memecah kalimat menjadi kata tunggal, dilanjutkan dengan stopword removal untuk menghapus kata umum yang tidak memiliki kontribusi signifikan terhadap analisis. Tahap terakhir adalah stemming, yaitu mengubah kata menjadi bentuk dasar agar variasi kata dengan makna serupa dapat diproses secara lebih konsisten pada tahap klasifikasi.

Metode

- Data Labeling

Pelabelan dilakukan dengan menerapkan pendekatan berbasis lexicon-based, yaitu metode yang menentukan label sentimen berdasarkan keberadaan kata-kata yang terdapat dalam kamus sentimen. Setiap komentar dianalisis dengan membandingkan jumlah kata yang memiliki polaritas positif dan negative. Komentar diberi label positif jika ada lebih banyak kata positif dari pada kata negatif, dan komentar diberi label negatif jika ada lebih banyak kata negatif dari pada positif. Jika jumlah keduanya sama atau tidak ditemukan kata yang terdapat pada kamus sentimen, maka komentar diberi label netral. Untuk meningkatkan kualitas pelabelan, hasil pelabelan otomatis selanjutnya divalidasi secara manual.

- TF-IDF

TF-IDF merupakan kombinasi antara frekuensi kemunculan kata dalam satu dokumen dan tingkat keunikan kata tersebut dalam keseluruhan dataset. Melalui pembobotan ini, kata yang sering muncul namun umum akan memiliki bobot rendah, sedangkan kata yang jarang namun spesifil

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

N = jumlah dokumen
Dfi = jumlah kemunculan term terhadap i
Tfij = jumlah kemunculan i di j

Metode

- Support Vector Machine (SVM) digunakan sebagai metode klasifikasi sentimen dalam penelitian ini. Secara prinsip, SVM bekerja dengan mencari *hyperplane* atau batas pemisah optimal yang mampu membedakan data berdasarkan pola yang dimilikinya. Pada analisis sentimen, hyperplane tersebut memisahkan komentar ke dalam kategori positif dan negatif. Jika data bersifat linier, digunakan model linier, sedangkan untuk pola yang lebih kompleks digunakan fungsi kernel untuk memetakan data ke ruang fitur yang lebih tinggi. SVM dipilih karena memiliki kemampuan generalisasi yang baik dan stabil pada data berdimensi tinggi seperti teks.

$$f(x) = sign\ (w.\ x + b)$$

Rumus Kernel	
Kernel Linear	$K(x, y) = (x.\ y)$
Kernel Polinominal	$K(x, y) = (x, y + c)$d
Kernel RBF	$K(x, x') = exp\ (-\gamma(x, y)^2$

$f(x)$ = Fungsi Prediksi
w = Vektor Normal Hyperplane
x = Vektor Input
b = Bias
c = Coefficient
d = Derajat Polinominal

Metode

- Evaluasi Model

Tahap evaluasi dilakukan untuk menilai kinerja model SVM dalam mengklasifikasikan sentimen. Pada penelitian ini digunakan metrik akurasi, yaitu perbandingan antara jumlah prediksi yang benar dengan total seluruh data uji. Nilai akurasi menunjukkan seberapa tepat model dalam mengelompokkan komentar ke dalam kelas sentimen yang sesuai.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Jumlah\ Data}$$

- Visualisasi Data

Visualisasi data pada penelitian ini digunakan untuk memudahkan interpretasi hasil klasifikasi sentimen yang diperoleh dari model Support Vector Machine (SVM). Visualisasi dilakukan menggunakan confusion matrix yang menyajikan perbandingan antara label aktual dengan label hasil prediksi model. Dari confusion matrix dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi pada setiap kategori sentimen.

Hasil dan Pembahasan

- Preprocessing

Text	Cleaned	Stopwords Removed	Tokens	Stemmed	Final Text
tim yg suka jnt ketimbang spx JNT sat set spx lelet	tim yg suka jnt ketimbang spx jnt sat set spx lelet	tim suka jnt ketimbang spx jnt sat set spx lelet	tim,suka,jnt, ketimbang,spx,jnt,sat,set, spx,lelet	tim,suka, jnt,ketimbang,spx, jnt,sat,set ,spx,lelet	tim suka jnt ketimbang spx jnt sat set spx lelet
yg harus di stop itu spx 😠😠 subuh menuju alamat pengiriman nyampenya magrib	yg harus di stop itu spx subuh menuju alamat pengiriman nyampenya magrib	stop spx subuh menuju alamat pengiriman nyampenya magrib	stop,spx,subuh,menuju,alamat,pengiriman,nyampenya,magrib	stop,spx, subuh,tuju,alamat, kirim,nyampenya, magrib	stop spx subuh tuju alamat kirim nyampenya magrib

Proses diawali dengan cleaning untuk menghilangkan elemen-elemen yang tidak diperlukan. Setelah itu dilakukan stopword removal guna menghapus kata-kata yang sering muncul tetapi tidak memberikan informasi penting terhadap sentimen komentar. Tahap selanjutnya adalah tokenizing, yaitu memecah kalimat menjadi beberapa token atau kata tunggal sehingga setiap kata dapat diproses secara terpisah. Setelah proses tokenisasi, dilakukan stemming untuk mengubah kata menjadi bentuk dasarnya. Proses ini bertujuan untuk menyamakan variasi kata yang memiliki makna serupa sehingga dapat mengurangi keragaman istilah dalam dataset. Sebagai contoh, kata "pengiriman" akan direduksi menjadi "kirim", sedangkan kata "berbelanja" menjadi "belanja".

Hasil dan Pembahasan

- Data Labeling

No	Komentar (Final Text)	Label Sentimen
1	padahal banding ekspedisi ny jnt paling recommended gercep sat set sesuai estimasi malah kadang lebih cepat sampai dari perkiraan	Positif
2	bagus jnt cepat nyampek	Positif
3	jnt terlalu gercep jadi banned Shopee	Positif
4	spx lelet kembali jnt dulu	Positif
5	jnt kurir area kurir spx tiap antar barang beda orang	Netral

Diperoleh dataset akhir sebanyak 5.697 komentar yang telah melalui proses pelabelan menggunakan pendekatan lexicon-based dan validasi manual. Hasil pelabelan menunjukkan bahwa terdapat 1.899 komentar bersentimen positif, 1.899 komentar bersentimen negatif, dan 1.899 komentar bersentimen netral. Dataset hasil pelabelan tersebut selanjutnya digunakan sebagai data masukan pada tahap ekstraksi fitur menggunakan metode TF-IDF sebelum dilakukan proses klasifikasi menggunakan Support Vector Machine (SVM).

Hasil dan Pembahasan

- TF-IDF

No	Kata	Mean TF-IDF
1	jnt	0,048613
2	spx	0,034816
3	kurir	0,033074
4	paket	0,026559
5	lebih	0,020287
6	shopee	0,020126
7	kirim	0,017900
8	bener	0,017685
9	tiktok	0,017019
10	ga	0,016579

Berdasarkan hasil perhitungan Mean TF-IDF, kata jnt memiliki nilai bobot rata-rata tertinggi sebesar 0,048613, diikuti oleh kata spx sebesar 0,034816 dan kurir sebesar 0,033074. Tingginya nilai bobot pada ketiga kata tersebut menunjukkan bahwa pembahasan mengenai J&T Express, SPX, dan layanan kurir menjadi topik yang paling dominan dalam komentar pengguna. Selain itu, kata-kata seperti paket, shopee, kirim, lebih, bener, dan tiktok juga memiliki nilai bobot TF-IDF yang relatif tinggi.

Hasil dan Pembahasan

- Klasifikasi Suport Vector Machine (SVM)

Data yang telah melalui proses ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) selanjutnya dibagi menjadi data latih (training data) dan data uji (testing data). Pembagian data dilakukan menggunakan metode train-test split dengan rasio 80:20. Dari total 5.697 data komentar yang digunakan dalam penelitian, sebanyak 4.557 data digunakan sebagai data latih dan 1.140 data digunakan sebagai data uji. Pembagian data ini bertujuan agar model memperoleh data yang cukup untuk proses pembelajaran sekaligus dapat dievaluasi menggunakan data yang belum pernah digunakan selama proses pelatihan.

Setelah proses pembagian data selesai, tahap berikutnya adalah klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM). Algoritma SVM digunakan untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral berdasarkan fitur hasil pembobotan TF-IDF. Selanjutnya, kinerja model dievaluasi menggunakan data uji untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen secara tepat.

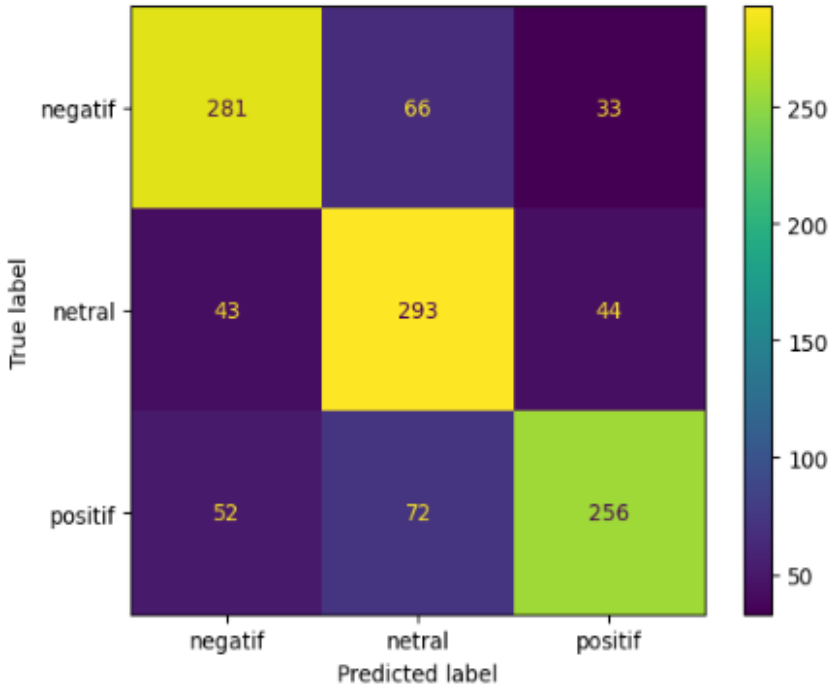
Hasil dan Pembahasan

- Evaluasi Model

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0,86	0,76	0,80	380
Netral	0,68	0,87	0,77	380
Positif	0,83	0,70	0,76	380
Accuracy	-	-	77,54%	1.14
Macro Average	0,79	0,78	0,78	1.14
Weighted Average	0,79	0,78	0,78	1.14

Hasil dan Pembahasan

- Confusion Matrix



Hasil confusion matrix menunjukkan bahwa sebagian besar komentar telah berhasil diklasifikasikan sesuai dengan label sebenarnya. Meskipun masih terdapat beberapa kesalahan klasifikasi pada setiap kategori sentimen, model Support Vector Machine (SVM) mampu mengklasifikasikan 830 dari 1.140 data uji dengan benar, sehingga menghasilkan tingkat akurasi sebesar 77,54%.

Kesimpulan

Berdasarkan hasil penelitian mengenai analisis sentimen publik di media sosial TikTok terhadap penghentian layanan J&T Express pada Shopee menggunakan metode Support Vector Machine (SVM), dapat disimpulkan bahwa isu penghentian kerja sama antara Shopee dan J&T Express memunculkan berbagai tanggapan dari pengguna TikTok yang dapat diklasifikasikan ke dalam sentimen positif, negatif, dan netral.

Hasil klasifikasi menggunakan kombinasi metode Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma Support Vector Machine (SVM) mendapatkan akurasi sebesar 77,54% pada data uji sebanyak 1.140 komentar. Berdasarkan hasil evaluasi menggunakan confusion matrix, model berhasil mengklasifikasikan 830 komentar dengan benar yang terdiri atas 281 komentar negatif, 293 komentar netral, dan 256 komentar positif. Hasil tersebut menunjukkan bahwa kombinasi metode TF-IDF dan SVM mampu digunakan untuk mengklasifikasikan sentimen publik terhadap isu penghentian kerja sama antara Shopee dan J&T Express dengan tingkat ketepatan yang cukup baik.

Universitas
Muhammadiyah
Sidoarjo

Terimakasih