Perpustakaan Umsida

TA_Fathur Rahman Maulana_cpl.docx

ANITA

Kampus 3

Perpustakaan

Document Details

Submission ID

trn:oid:::1:3619564038

Submission Date

Jul 30, 2026, 9:29 AM GMT+7

Download Date

Jul 30, 2026, 9:50 AM GMT+7

File Name

TA_Fathur_Rahman_Maulana_cpl.docx

File Size

414.2 KB

10 Pages

4,334 Words

28,593 Characters

19% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Small Matches (less than 10 words)

Top Sources

13% Internet sources
11% Publications
13% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

13% Internet sources
11% Publications
13% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	Universitas Pancasila 1%
2	Student papers	UPN Veteran Yogyakarta 1%
3	Internet	journal.stiestekom.ac.id 1%
4	Internet	repository.upnvj.ac.id <1%
5	Student papers	Universitas Putera Batam <1%
6	Student papers	Universitas Sangga Buana YPKP <1%
7	Internet	kc.umn.ac.id <1%
8	Publication	Al Khaidar Author. "ANALISIS SENTIMEN DI INSTAGRAM TERHADAP MENTERI KEU... <1%
9	Student papers	Universitas Bengkulu <1%
10	Internet	ejurnal.seminar-id.com <1%
11	Publication	Ayu Oktaviyanti. "ANALISIS SENTIMEN ULASAN PELECEHAN DI MEDIA SOSIAL TIK... <1%

12	Internet	
ejurnal.stmik-budidarma.ac.id		<1%
13	Publication	
Evip Athirah Salsabila, Mellysa Pratama, Putri Wahyuni, Vina Purwaningrum, Mufi...		<1%
14	Student papers	
Universitas Islam Indonesia		<1%
15	Student papers	
Universitas Muhammadiyah Sidoarjo		<1%
16	Internet	
jurnal.polibatam.ac.id		<1%
17	Internet	
hostjournals.com		<1%
18	Internet	
journal.arteii.or.id		<1%
19	Internet	
pssh.umsida.ac.id		<1%
20	Publication	
Arimbi Puspitasari, Diana Sava Salsabila, Dwi Roliawati. "Penerapan ResNet-50 C...		<1%
21	Internet	
ejournal.itn.ac.id		<1%
22	Internet	
archive.umsida.ac.id		<1%
23	Internet	
eprints2.undip.ac.id		<1%
24	Publication	
Eko Fuji Pangestu -, Bambang Irawan -. "KLASIFIKASI PENYAKIT TANAMAN PADI ...		<1%
25	Student papers	
Universitas Islam Riau		<1%

26 Student papers
Universitas Kristen Wira Wacana Sumba <1%

27 Internet
www.coursehero.com <1%

28 Publication
Ghefira Zahra Nur Fadhilah, Muh. Yamin, Asa Hari Wibowo. "Sistem Cerdas Dete... <1%

29 Publication
Indah Sartika, Imam Saputra. "Analisis Sentimen Masyarakat Terhadap Pembatas... <1%

30 Publication
Paulina Gorat Frans, Frans Steven Pakpahan, Sardo Pardingotan Sipayung. "Anal... <1%

31 Student papers
Universitas Pamulang <1%

32 Student papers
Universitas Tarumanagara <1%

33 Student papers
Universiti Tun Hussein Onn Malaysia <1%

34 Internet
jurnalsyntaxadmiration.com <1%

35 Internet
lmj.areeo.ac.ir <1%

36 Internet
prosiding.aritekin.or.id <1%

37 Internet
hargahp.co.id <1%

38 Internet
journal.unismuh.ac.id <1%

39 Internet
languar.net <1%

40 Publication

AHMAD ABBDILLAH HISYAM -. "ANALISIS SENTIMEN PERSEPSI PUBLIK TERHADAP ... <1%

41 Publication

M. Saleh, Dwi Yuli Prasetyo. "Analisis Sentimen Konsumen Terhadap Usaha Food ... <1%

42 Publication

Nova Tri Romadloni, Pamela Hana Mulia, Wiwit Supriyanti. "Perbandingan Algorit... <1%

43 Internet

diajeng.lldikti6.id <1%

44 Internet

e-jurnal.lppmunsera.org <1%

45 Internet

repository.uinjkt.ac.id <1%

46 Internet

repository.unissula.ac.id <1%

Analisis Sentimen Publik di Media Sosial TikTok terhadap Penghentian Layanan J&T Express pada Shopee Menggunakan Metode Support Vector Machine (SVM)

Public Sentiment Analysis on TikTok Social Media Regarding the Termination of J&T Express Services on Shopee Using the Support Vector Machine (SVM) Method

Fathur Rahman Maulana, Istian Kriya Almanfaluti, S.Kom., M.Kom., Bayu Hari Prasojo, S.Si., M.Pd. Andry Rachmadany, S.Kom., M.Kom.
[1)] Program Studi Bisnis Digital, Universitas Muhammadiyah Sidoarjo,
*Email Penulis Korespondensi: istian.alman@umsida.ac.id

Abstract. *The termination of the partnership between Shopee and J&T Express has become a widely discussed issue among social media users, particularly on TikTok. The various comments reflect the diverse public responses to the change in logistics services. This study aims to analyze public sentiment regarding the termination of J&T Express services on Shopee using the Support Vector Machine (SVM) method. The research data was obtained through a scraping process of 5,760 comments. After preprocessing, 5,697 comments were obtained, which were used in the analysis process. The sentiment labeling process was carried out using a lexicon-based approach and manually validated to improve labeling quality. Next, word weighting was carried out using Term Frequency–Inverse Document Frequency (TF-IDF), dividing the data into 80% training data and 20% test data, and then the classification process using the Support Vector Machine (SVM) algorithm. The test results showed that the SVM model achieved an accuracy of 77,54% and was able to classify public sentiment into positive, negative, and neutral categories. The results of the study show that the combination of the TF-IDF and Support Vector Machine (SVM) methods can be used to analyze public sentiment towards the issue of terminating cooperation between Shopee and J&T Express with a fairly good level of accuracy.*

Keywords - *Sentiment Analysis, TikTok, Shopee, J&T Express, TF-IDF, Support Vector Machine.*

Abstrak. Pemutusan kerja sama antara Shopee dan J&T Express menjadi salah satu isu yang banyak dibahas oleh pengguna media sosial, khususnya di TikTok. Berbagai komentar yang muncul mencerminkan beragam tanggapan masyarakat terhadap perubahan layanan logistik tersebut. Penelitian ini memiliki tujuan menganalisis sentimen publik terhadap penghentian layanan J&T Express pada Shopee menggunakan metode Support Vector Machine (SVM). Data penelitian diperoleh melalui proses scraping sebanyak 5.760 komentar. Setelah dilakukan preprocessing, diperoleh 5.697 komentar yang digunakan dalam proses analisis. Proses pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based dan divalidasi secara manual untuk meningkatkan kualitas pelabelan. Selanjutnya dilakukan pembobotan kata menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), pembagian data menjadi 80% data latih dan 20% data uji, kemudian proses klasifikasikan menggunakan algoritma Support Vector Machine (SVM). Hasil pengujian menunjukkan bahwa model SVM memperoleh akurasi sebesar 77,54% serta mampu mengklasifikasikan sentimen publik ke dalam kategori positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa kombinasi metode TF-IDF dan Support Vector Machine (SVM) dapat digunakan untuk menganalisis sentimen publik terhadap isu penghentian kerja sama antara Shopee dan J&T Express dengan tingkat ketepatan yang cukup baik.

Kata Kunci - Analisis Sentimen, TikTok, Shopee, J&T Express, TF-IDF, Support Vector Machine.

I. PENDAHULUAN

Kemajuan teknologi informasi dan digital belakangan ini berkembang sangat pesat serta membawa banyak perubahan dalam kehidupan sehari-hari. Salah satu yang paling terasa adalah perubahan cara orang berkomunikasi, berdagang, hingga berbelanja [1]. Masyarakat kini lebih memilih cara yang praktis dan cepat, termasuk dalam melakukan transaksi jual beli secara online melalui e-commerce [2]. Hal ini menjadi daya tarik bagi masyarakat karena kini aktivitas belanja tidak harus dilakukan secara langsung dengan datang ke toko fisik, melainkan dapat

dilakukan hanya dengan melalui perangkat digital yang dapat di akses dimana saja dan kapan saja [3]. Perkembangan infrastruktur digital seperti akses internet yang semakin mudah juga menjadi salah satu faktor pendorong minat masyarakat Indonesia terhadap belanja online [4].

Meningkatnya minat masyarakat untuk berbelanja melalui toko online mendorong banyak platform e-commerce untuk terus meningkatkan kualitas layanan dan fitur pada aplikasi mereka. Persaingan pada industry e-commerce yang semakin ketat yang menuntut platformnya untuk bisa memberikan pengalaman pengguna yang lebih baik dan efisien [5]. Salah satu platform dengan tingkat kunjungan dan popularitas tertinggi di Indonesia adalah Shopee. Shopee merupakan salah satu marketplace e-commerce yang memiliki jumlah pengguna cukup besar dan diminati oleh masyarakat Indonesia. Shopee merupakan platform belanja online yang berasal dari Singapura dan berada di bawah naungan perusahaan teknologi Sea Group. Aplikasi ini pertama kali diluncurkan di Singapura pada tahun 2015 dan pada tahun yang sama mulai memasuki pasar Indonesia [6].

Hingga tahun 2025, Shopee masih menunjukkan dominasinya di pasar e-commerce nasional dan menjadi salah satu aplikasi belanja online yang paling banyak digunakan oleh masyarakat Indonesia. Beberapa penelitian juga menunjukkan bahwa Shopee menjadi pilihan utama konsumen karena kemudahan penggunaan, kepercayaan terhadap platform, serta promosi yang menarik dibandingkan dengan kompetitornya [7]. Berdasarkan data persaingan toko online di Indonesia tahun 2019, Shopee menempati posisi kedua di bawah Tokopedia dalam jumlah pengunjung dan pengikut media sosial. Namun, dari tahun ke tahun jumlah pengguna Shopee terus meningkat hingga akhirnya mampu menjadi salah satu platform e- commerce dengan dominasi tertinggi di Indonesia pada 2025. Hal ini menunjukkan bahwa meskipun sempat berada di posisi kedua, Shopee berhasil memperkuat dominasinya di pasar e-commerce nasional. Tingginya jumlah pengguna tersebut juga menuntut perusahaan untuk terus menjaga kualitas layanan agar tetap kompetitif dan mampu mempertahankan posisinya sebagai salah satu aplikasi belanja online terpopuler di kawasan Asia Tenggara [8].

Untuk mepertahankan reputasinya sebagai salah satu platform e-commerce terpopuler, Shopee berupaya untuk terus berinovasi di tengah tren perkembangan digital [9]. Shopee tidak hanya fokus pada peningkatan fitur teknis seperti antarmuka pengguna dan pembaruan sistem, tetapi juga secara konsisten menjaga kualitas layanan demi memberikan pengalaman pengguna yang optimal [10]. Beberapa langkah yang juga dilakukan oleh perusahaan seperti peningkatan keamanan transaksi, penerapan pengawasan sistem terhadap mitra penjual, serta optimalisasi tampilan aplikasi agar lebih ramah pengguna [11]. Kualitas layanan elektronik atau e-service quality juga menjadi faktor penting yang memengaruhi tingkat kepuasan pengguna dalam menggunakan platform Shopee [12]. Langkah-langkah ini menjadi strategi berkelanjutan dalam menghadapi persaingan pasar digital yang semakin ramai dan menuntut setiap platform untuk terus beradaptasi, sekaligus menunjukkan komitmen Shopee dalam menjaga kepuasan dan loyalitas pelanggannya [13].

Belakangan ini, pengguna Shopee dihebohkan dengan isu pemutusan hubungan kerja sama antara Shopee dan salah satu mitra logistik utamanya, yaitu J&T Express. Kabar ini memicu kekhawatiran dari berbagai pihak, mengingat J&T telah menjadi layanan pengiriman andalan bagi banyak pengguna karena kecepatan dan jangkauan wilayahnya yang luas [14]. Dalam layanan logistik e-commerce, kualitas dan keandalan layanan pengiriman menjadi faktor penting yang memengaruhi persepsi pengguna terhadap kinerja sistem distribusi suatu platform [15]. Walaupun layanan J&T Cargo masih tersedia untuk pengiriman dalam skala besar, hal ini tentunya tetap menimbulkan pertanyaan mengenai keberlanjutan terkait efektivitas sistem logistik Shopee, serta potensi dampaknya terhadap kenyamanan dan kepuasan pengguna [16]. Penghentian kerja sama ini tentunya turut memengaruhi opsi jasa pengiriman yang tersedia saat berbelanja. Sehingga konsumen yang seringkali menggunakan J&T sebagai opsi pengirimannya kini perlu menyesuaikan diri dengan alternatif lain seperti Shopee Xpress, SiCepat, atau AnterAja. Sementara itu, bagi para penjual terutama yang berada di wilayah terpencil dan sebelumnya sangat bergantung pada J&T perubahan ini dapat menimbulkan kendala logistik baru dalam proses pemenuhan pesanan [17].

Banyak penelitian yang melakukan studi mengenai sentimen pengguna terhadap e-commerce, tetapi sebagian besar hanya berfokus pada topik terkait kualitas layanan Shopee dan kepuasan pelanggan. penelitian terdahulu lebih banyak menggunakan data dari Twitter, Google Play Store, atau Instagram, sehingga analisis berbasis komentar TikTok masih jarang dilakukan. Hingga saat ini masih belum ditemukan penelitian yang secara spesifik menganalisis pemutusan kerjasama antara Shopee dan J&T Express. Selain itu, belum ada penelitian yang memadukan isu pemutusan mitra logistik dengan data TikTok serta metode klasifikasi SVM sehingga topik tersebut menunjukkan bahwa terdapat celah analisis yang belum dikaji lebih mendalam.

Maka berdasarkan hal tersebut, dalam penelitian ini akan membahas mengenai sentimen publik terhadap isu pemutusan kerja sama antara Shopee dan J&T Express. Penelitian ini memiliki tujuan untuk mengetahui bagaimana tanggapan masyarakat, khususnya pengguna media sosial TikTok, terhadap kebijakan tersebut [18]. Untuk memperoleh data, peneliti melakukan metode scraping pada platform TikTok untuk mengumpulkan opini dan reaksi pengguna [19]. Kemudian, data yang didapatkan selanjutnya akan dianalisis menggunakan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi guna mengidentifikasi serta mengkategorikan sentimen menjadi positif,

negatif, atau netral [20]. Melalui pendekatan ini, diharapkan penelitian dapat memberikan gambaran yang objektif mengenai persepsi publik serta implikasinya terhadap reputasi dan citra Shopee di mata konsumen.

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.

II. LITERATUR REVIEW

A. E-Commerce

E-commerce merupakan aktivitas jual beli barang, layanan, maupun pertukaran informasi yang dilakukan melalui jaringan komputer seperti internet. Dalam praktiknya, model perdagangan ini memungkinkan penjual dan pembeli untuk berinteraksi secara online sehingga biaya operasional dapat ditekan, proses transaksi menjadi lebih efisien, dan produktivitas meningkat. Selain itu, E-commerce juga mendukung pertukaran data secara digital, seperti penggunaan email, pembayaran elektronik, dan dokumen tanpa kertas [21]. E-commerce sangat disukai masyarakat karena aksesnya yang mudah, banyak pilihan produk, dan fitur-fitur yang mempermudah transaksi. Platform marketplace seperti Shopee bukan hanya tempat jualan, tapi juga ekosistem yang menghubungkan penjual, pembeli, sistem pembayaran, dan layanan logistik. Karena itu, performa sebuah e-commerce seringkali diukur dari pengalaman pengguna mulai dari proses pencarian produk, checkout, sampai layanan purna jual.

B. Mitra Logistik

Di Indonesia, kebutuhan distribusi yang semakin cepat dan efisien membuat manajemen jasa transportasi dan logistik berkembang sangat pesat. Industri ini dituntut untuk mampu menyediakan layanan yang andal, pengiriman yang tepat waktu, serta akses layanan yang mudah bagi pelanggan, karena keberhasilan sebuah perusahaan logistik bergantung pada bagaimana mereka mengelola keandalan layanan, ketepatan pengiriman, dan kemudahan sistem pelacakan bagi pengguna [22]. Dalam konteks e-commerce, 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. Selain itu, perubahan kerja sama dengan mitra logistik juga akan mempengaruhi stabilitas layanan, khusunya pada wilayah yang bergantung pada satu penyedia jasa logistik.

C. Sentiment Publik

Analisis sentimen merupakan salah satu 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 [23]. Saat ini, informasi berbasis teks banyak tersebar di forum, blog, dan terutama media sosial, sehingga platform seperti TikTok menjadi sumber data yang sangat kaya untuk melihat bagaimana publik menilai suatu isu. TikTok menghasilkan konten singkat yang cepat viral dengan kolom komentar yang dipenuhi opini spontan, sehingga relevan untuk dianalisis. Namun begitu, karakter bahasa di TikTok yang sering menggunakan slang, emoji, dan campuran bahasa membuat proses pemrosesan teks membutuhkan penanganan khusus agar sistem dapat membaca sentimen dengan lebih akurat.

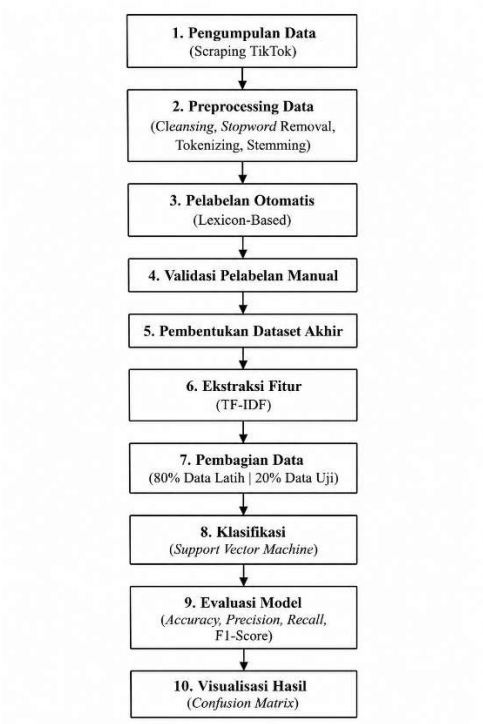
D. Suport Vector Machine

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. Dalam proses penerapannya, data teks harus diubah menjadi bentuk vektor terlebih dahulu, yang kemudian diberi bobot menggunakan TF-IDF agar fitur-fitur yang dianggap penting bisa muncul lebih dominan. Kekuatan utama SVM

terletak pada kemampuannya mencari batas pemisah (hyperplane) yang paling optimal sehingga masing-masing kelas dapat dipisahkan dengan margin yang lebih lebar. Akan tetapi, metode ini tetap memiliki keterbatasan seperti, ketika sejumlah fitur memiliki nilai yang terlalu mirip, hal tersebut bisa menurunkan akurasi. Selain itu, SVM memerlukan proses pelatihan yang bertahap serta penyesuaian parameter agar hasil klasifikasi benar-benar maksimal, terutama pada data berbahasa Indonesia seperti komentar TikTok yang cenderung variatif dan tidak baku [24].

III. METODE

Penelitian ini memakai metode *Support Vector Machine* (SVM) sebagai teknik utama dalam proses klasifikasinya. Secara umum, ada tujuh tahapan yang dilakukan. Proses dimulai dari pengumpulan data dengan teknik *crawling*, lalu data yang terkumpul masuk ke tahap pra-pemrosesan. Setelah itu, setiap data diberi label untuk membedakan komentar yang bersentimen positif dan negatif. Data yang sudah dilabeli kemudian diubah ke bentuk numerik melalui proses pembobotan kata menggunakan TF-IDF, sebelum akhirnya diklasifikasikan memakai SVM. Pada bagian akhir, performa model dievaluasi dengan *stratified k-fold cross validation*, kemudian hasilnya divisualisasikan [25].



A. 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. Pengumpulan data dilakukan menggunakan pemrograman Python untuk mengekstraksi komentar pengguna secara otomatis. Dari proses tersebut, diperoleh sebanyak 5.760 komentar yang berisi berbagai opini dan sentimen warganet, dengan data disimpan dalam format file CSV sebagai bahan analisis selanjutnya. Penggunaan jumlah data tersebut diperkuat dengan pendapat [26] yang menyatakan bahwa dalam penelitian berbasis big data dan media sosial, representativitas data tidak semata-mata ditentukan oleh penggunaan seluruh populasi data, melainkan oleh kemampuan data tersebut dalam merepresentasikan fenomena sosial yang diteliti.

Pembagian *dataset* pada penelitian ini dilakukan menggunakan pendekatan *holdout method*, yaitu dengan memisahkan data ke dalam *training set* dan *testing set.* Metode ini mempartisi data secara acak menjadi dua himpunan independen, yang mana sebagian data digunakan untuk proses pelatihan model, sedangkan sisanya dipergunakan untuk pengujian kinerja model terhadap data yang tidak terlibat dalam pelatihan [27]. Proporsi 80% data latih dan 20% data uji dipilih untuk memperoleh keseimbangan antara jumlah data yang memadai dalam proses

turnitin Page 11 of 16 - Integrity Submission Submission ID trn:oid:::1:3619564038

pembelajaran model dan keandalan evaluasi performa. Pemisahan data pelatihan dan data pengujian tersebut merupakan prosedur yang lazim digunakan dalam holdout method guna mengukur kemampuan generalisasi model secara objektif [28].

B. Pra-Pemrosesan Data

Pada tahap *preprocessing*, data komentar yang diperoleh terlebih dahulu melalui proses cleansing untuk menghapus berbagai elemen yang dianggap sebagai *noise*, seperti *username*, karakter HTML, emoji, hashtag, URL, hingga simbol-simbol lain yang tidak relevan untuk analisis. Setelah itu dilakukan *case folding* dengan menyeragamkan seluruh teks ke huruf kecil serta menghilangkan karakter selain huruf yang tidak diperlukan. Komentar yang sudah bersih kemudian diproses pada tahap *tokenizing*, yaitu memecah kalimat menjadi potongan kata tunggal dengan memperhatikan batas spasi dan tanda baca. Selanjutnya dilakukan *stopword removal* atau *filtering* untuk menghapus kata-kata umum yang tidak memiliki kontribusi penting terhadap analisis, seperti kata sambung dan istilah yang terlalu generik. Tahap terakhir merupakan stemming, yaitu proses merubah kata menjadi bentuk dasar agar variasi kata yang mempunyai makna serupa dapat diproses secara lebih konsisten pada tahap klasifikasi [29].

C. Data Labeling

Setelah proses prapemrosesan selesai, tahap berikutnya adalah pemberian label pada setiap komentar untuk menentukan arah sentimennya. 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 [30]. Untuk meningkatkan kualitas pelabelan, hasil pelabelan otomatis selanjutnya divalidasi secara manual. Proses validasi dilakukan dengan meninjau kembali hasil pelabelan sehingga label yang kurang sesuai dapat diperbaiki.

D. TF-IDF

Pada tahap selanjutnya, setiap kata dalam komentar yang sudah diproses akan diberi bobot menggunakan metode TF–IDF. Pembobotan ini dilakukan untuk melihat seberapa penting suatu kata dalam satu komentar dibandingkan dengan kumpulan data secara keseluruhan. Nilai *term frequency* menunjukkan seberapa sering kata tersebut muncul dalam satu dokumen, sedangkan *inverse document frequency* menunjukkan seberapa umum atau jarangnya kata tersebut muncul di seluruh dokumen. [31]. Melalui kombinasi kedua komponen ini, TF–IDF membantu menonjolkan kata-kata yang benar-benar memiliki nilai informasi sehingga dapat memperkuat proses analisis sentimen sebelum masuk ke tahap klasifikasi.

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

w_{ij} = bobot kata i pada dokumen j

N = jumlah dokumen

df_i = jumlah kemunculan term terhadap i

tf_{ij} = jumlah kemunculan i di j

E. Klasifikasi Metode SVM

Support Vector Machine (SVM) digunakan dalam penelitian ini sebagai metode utama untuk melakukan klasifikasi sentimen. Secara prinsip, algoritma ini bekerja dengan mencari sebuah garis pemisah atau *hyperplane* yang mampu membedakan data berdasarkan pola yang dimilikinya. Pada kasus analisis sentimen, *hyperplane* tersebut berfungsi memisahkan komentar dengan sentimen positif, negatif, maupun netral. Jika data dapat dipisahkan dengan mudah menggunakan garis lurus, maka SVM cukup menggunakan model linier. Namun ketika pola data tidak bersifat

turnitin Page 11 of 16 - Integrity Submission Submission ID trn:oid:::1:3619564038

linier, dengan menggunakan fungsi kernel, SVM memetakan data ke ruang fitur yang lebih tinggi sehingga membuat pembentukan batas pemisah menjadi lebih jelas. Penggunaan kernel seperti kernel linier, polinomial, maupun gaussian membantu SVM menangani pola yang kompleks dan tidak teratur. Keunggulan utama SVM adalah kemampuannya dalam melakukan generalisasi dengan baik, stabil pada data berdimensi tinggi, serta menghasilkan akurasi yang kompetitif pada tugas klasifikasi teks [29]. Berikut rumus umum untuk metode SVM:

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

Algoritma Support Vector Machine (SVM) pada penelitian ini menerapkan kernel linear sebagai fungsi kernel dalam proses klasifikasi. Pemilihan kernel tersebut disesuaikan dengan karakteristik data teks yang telah dikonversi menjadi representasi numerik melalui metode TF-IDF. Dengan pendekatan tersebut, model membangun hyperplane yang digunakan untuk memisahkan data berdasarkan kelas sentimennya. Rumus fungsi kernel linear adalah sebagai berikut:

$$K(x, y) = (x.\,y)$$

$f(x)$ = fungsi prediksi

K(x, y) = fungsi kernel linear yang digunakan untuk mengukur kemiripan antara dua vektor data

w = vektor normal hyperplane

x = vektor input

y = vektor input

b = bias

F. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai seberapa baik model SVM bekerja dalam mengklasifikasikan sentimen. Proses ini penting karena sebuah model klasifikasi perlu diukur performanya agar diketahui tingkat ketepatan prediksi yang dihasilkan.

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

Dalam penelitian ini, penilaian dilakukan menggunakan metrik akurasi, yaitu persentase prediksi yang sesuai dengan label sebenarnya. Nilai akurasi memberikan gambaran mengenai kemampuan model dalam mengenali dan mengelompokkan data ke kelas sentimen yang tepat secara keseluruhan [32]. Dengan demikian, evaluasi ini berfungsi untuk memastikan bahwa model tidak hanya mampu belajar dari data, tetapi juga memberikan hasil klasifikasi yang reliabel pada data uji.

G. 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. Sehingga visualisasi ini memudahkan dalam identifikasi jumlah data yang berhasil diprediksi dengan benar maupun data yang masih mengalami kesalahan klasifikasi pada setiap kelas sentimen. Oleh karena itu, confusion matrix digunakan untuk membantu menganalisis kinerja model klasifikasi yang telah dibangun [33].

IV. HASIL DAN PEMBAHASAN

A. Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum memasuki proses analisis sentimen. Pada tahap ini dilakukan beberapa proses, yaitu cleaning, stopword removal, tokenizing, dan stemming.

Tabel 1 *Preprocessing Data*

Berdasarkan Tabel 1, data komentar yang diperoleh dari proses scraping masih mengandung berbagai elemen yang dapat mengganggu proses analisis, seperti emoji, tanda baca, URL, karakter khusus, serta variasi penulisan kata. Oleh karena itu, dilakukan tahap preprocessing untuk menghasilkan data yang lebih bersih dan siap digunakan pada proses analisis.

Proses diawali dengan cleaning untuk menghilangkan elemen-elemen yang tidak diperlukan. Selanjutnya dilakukan stopword removal guna menghapus kata-kata yang sering muncul tetapi tidak memberikan informasi penting terhadap sentimen komentar. Tahap berikutnya 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 setiap kata ke bentuk dasarnya. Proses ini bertujuan untuk menyamakan variasi kata yang memiliki makna serupa sehingga dapat mengurangi keragaman istilah dalam dataset. Sebagai contoh, kata "pengiriman" direduksi menjadi "kirim", sedangkan kata "berbelanja" direduksi menjadi "belanja".

Tabel 2 *Hasil Pemeriksaan Data Setelah Tahap Preprocessing*

Hasil akhir dari seluruh tahapan preprocessing disajikan pada kolom final_text. Setelah dilakukan pemeriksaan terhadap hasil preprocessing, ditemukan 63 data yang tidak memiliki nilai pada kolom final_text. Data tersebut tidak bisa digunakan dalam proses analisis sehingga dihapus dari dataset. Dengan demikian, jumlah data yang semula sebanyak 5.760 komentar menjadi 5.697 komentar yang selanjutnya digunakan pada tahap pelabelan sentimen, ekstraksi fitur menggunakan metode TF-IDF, dan proses klasifikasi menggunakan Support Vector Machine (SVM).

B. Data Labeling

Setelah melalui tahap preprocessing, sebanyak 5.697 komentar diberi label sentimen menggunakan pendekatan lexicon-based. Pelabelan dilakukan dengan mencocokkan kata-kata yang terdapat pada setiap komentar dengan kamus sentimen yang telah disusun sesuai dengan konteks penelitian. Apabila jumlah kata positif lebih banyak dibandingkan jumlah kata negatif, maka komentar diberi label positif. Sebaliknya, apabila jumlah kata negatif lebih banyak dibandingkan jumlah kata positif, maka komentar diberi label negatif. Sementara itu, komentar yang tidak menunjukkan dominasi kata positif maupun kata negatif diberi label netral. Untuk meningkatkan ketepatan hasil pelabelan, hasil pelabelan otomatis kemudian divalidasi secara manual dengan meninjau kembali kesesuaian label terhadap isi komentar. Proses validasi dilakukan untuk memperbaiki label yang kurang sesuai sehingga diperoleh dataset akhir yang digunakan pada tahap ekstraksi fitur menggunakan metode TF-IDF dan proses klasifikasi menggunakan Support Vector Machine (SVM).

Tabel 3 *Labeling Data*

Berdasarkan Tabel 3, 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).

C. TF-IDF

Setelah proses pelabelan selesai dilakukan, tahap berikutnya adalah ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF).

Tabel 4 *TF-IDF*

Berdasarkan hasil perhitungan Mean TF-IDF pada Tabel 4.4, 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.

Hal ini menunjukkan bahwa mayoritas diskusi pengguna berfokus pada proses pengiriman barang, layanan ekspedisi, serta tanggapan pengguna terhadap penghentian kerja sama antara Shopee dan J&T Express. Hasil pembobotan TF-IDF tersebut menunjukkan bahwa fitur yang dihasilkan mampu merepresentasikan informasi penting pada setiap dokumen sehingga dapat digunakan sebagai masukan pada proses klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM).

D. Klasifikasi Menggunakan Support 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.

E. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan f1-score.

Tabel 5

Berdasarkan hasil evaluasi pada Tabel 5, model Support Vector Machine (SVM) memperoleh akurasi sebesar 77,54% pada data uji sebanyak 1.140 komentar. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sentimen komentar ke dalam tiga kategori, yaitu positif, negatif, dan netral dengan tingkat ketepatan yang cukup baik.

Pada kelas negatif, model memperoleh nilai precision 0,86, recall 0,76, dan F1-score 0,80. Untuk kelas netral diperoleh precision 0,68, recall 0,87, dan F1-score 0,77. Nilai recall yang lebih tinggi pada kelas netral menunjukkan bahwa sebagian besar komentar netral berhasil dikenali oleh model, meskipun masih terdapat sejumlah komentar dari kelas lain yang diprediksi sebagai netral.

Sementara itu, pada kelas positif model menghasilkan precision 0,83, recall 0,70, dan F1-score 0,76. Hal ini menunjukkan bahwa prediksi komentar positif memiliki tingkat ketepatan yang cukup baik, namun masih terdapat beberapa komentar positif yang belum berhasil dikenali secara optimal. Secara keseluruhan, nilai macro average sebesar 0,79 untuk precision serta 0,78 untuk recall dan F1-score menunjukkan bahwa performa model cukup konsisten pada seluruh kelas sentimen. Nilai weighted average yang sama juga mengindikasikan bahwa model memiliki kinerja yang relatif seimbang karena jumlah data pada setiap kelas telah diseimbangkan melalui proses validasi pelabelan.

F. Confusion Matrix

***Gambar 1** Confusion Matrix*

Selain menggunakan metrik evaluasi, performa model juga dianalisis menggunakan confusion matrix sebagaimana disajikan pada Gambar 3. Visualisasi ini digunakan untuk memperlihatkan kesesuaian antara label aktual dan label hasil prediksi sehingga dapat diketahui jumlah prediksi yang tepat maupun kesalahan klasifikasi pada setiap kategori sentimen.

Berdasarkan Gambar 3, pada kelas negatif terdapat 281 komentar berhasil diklasifikasikan dengan benar sebagai sentimen negatif, sedangkan 66 komentar diprediksi sebagai netral dan 33 komentar diprediksi sebagai positif. Pada kelas netral, sebanyak 293 komentar berhasil diprediksi dengan benar, sementara 43 komentar diprediksi sebagai negatif dan 44 komentar diprediksi sebagai positif. Adapun pada kelas positif, sebanyak 256 komentar berhasil dikenali dengan benar, sedangkan 52 komentar diprediksi sebagai negatif dan 72 komentar diprediksi sebagai netral.

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%.

V. 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.