

Opini Publik Terhadap Kebijakan Tarif Impor Donald Trump Menggunakan Logistic Regression dan SMOTE

Dewa^{*}¹, Fujianto², Hafiz Irsyad³

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer Rekayasa, Universitas Multi Data Palembang

Email: ^{*}dewa_2226250101@mhs.mdp.ac.id, ²fujianto_2226250119@mhs.mdp.ac.id,

³hafizirsyad@mdp.ac.id

(Naskah masuk: 29 Mei 2025, diterima untuk diterbitkan: 31 Juli 2025)

Abstrak: Penelitian ini bertujuan untuk mengamati bagaimana opini publik bereaksi pada biaya masuk barang Donald Trump, yang ramai dibicarakan di dunia maya, terutama pada platform YouTube. Karena banyaknya pendapat yang muncul pada platform tersebut, penting sekali memakai cara yang cocok agar lebih paham akan opini yang disampaikan. Kami memakai metode logistic Regression yang digabung dengan metode SMOTE untuk penelitian ini, guna dalam menganalisis opini yang ada pada komentar yang tersebar luas pada platform YouTube. Kami membagi sentimen komentar jadi dua kelompok yaitu, positif atau negatif. Setelah dilakukan dari banyak tahapan analisis data, mulai dari membersihkan teks hingga menghapus karakter yang bukan termasuk teks menggunakan pre-processing dan ekstraksi fitur, Hasil yang telah dibangun dengan Logistic Regression, kemudian dataset akan diseimbangkan dengan SMOTE mendapatkan hasil precision pada sentimen negatif sebesar 95% dan sentimen positif sebesar 02%, recall pada sentimen negatif 74% dan sentimen positif 11%, F1-Score pada sentimen negatif 83% dan sentimen positif 03%, serta jumlah sample pada kelas sentimen negatif 210 dan kelas sentimen positif 9. dengan hasil penelitian ini Logistic Regression dan SMOTE dapat diterapkan.

Kata Kunci – Opini Publik, YouTube, Logistic Regression, SMOTE, Donald Trump

Abstract: This research aims to observe how the public reacts to Donald Trump's cost of entry for goods, which is widely discussed online, especially on the YouTube platform. Because of the many opinions that appear on the platform, it is important to use a suitable method to better understand the opinions expressed. We used the logistic regression method combined with the SMOTE method for this study, in order to analyze the opinions in the widespread comments on the YouTube platform. We divided the sentiment of comments into two groups, namely, positive or negative. After many stages of data analysis, starting from cleaning text to removing characters that are not included in the text using pre-processing and feature extraction, the results that have been built with Logistic Regression, then the dataset will be balanced with SMOTE to get the results of precision on negative sentiment of 95% and positive sentiment of 02%, recall on negative sentiment 74% and positive sentiment 11%, F1-Score on negative sentiment 83% and positive sentiment 03%, and the number of samples in the negative sentiment class 210 and positive sentiment class 9.

Keywords – Opini Public, YouTube, Regresi Logistik SMOTE, Donald Trump

1. PENDAHULUAN

Pada masa kini media sosial youtube menjadi tontonan yang banyak digemari oleh masyarakat yang dapat menyaksikan berbagai macam video serta juga dapat berbagi video yang diinginkan kapan saja dan dimana saja yang diunggah oleh pengguna lain di youtube. YouTube adalah sebuah situs web video sharing(berbagi video) yang populer dimana para pengguna dapat memuat, menonton, dan berbagi klip video secara gratis [1]. Munculnya media sosial seperti youtube ini melahirkan sisi positif dan negatif didalamnya. Sisi positifnya yaitu masyarakat bisa mendapatkan laju pertukaran informasi dan hiburan dengan cepat serta bermanfaat [2]. Penurunan tarif impor sebagai bentuk dari liberalisasi perdagangan suatu negara akan berdampak pada menurunnya produktivitas perusahaan dan menyebabkan perusahaan keluar dari pasar [3]. Kebijakan tarif impor yang diterapkan oleh Presiden Donald Trump selama masa pemerintahannya menjadi salah satu kebijakan ekonomi paling kontroversial dalam beberapa dekade terakhir. Kebijakan tersebut bertujuan untuk melindungi industri

domestik Amerika Serikat dengan memberlakukan tarif tinggi terhadap produk impor, terutama dari negara Tiongkok dan beberapa negara lain. Di YouTube, masyarakat tidak hanya bisa menonton berita atau konten terkait kebijakan tarif impor Donald Trump, tetapi juga dapat aktif dalam memberikan komentar, membuat video reaksi, hingga berdiskusi dalam berbagai video dan live streaming secara bebas dan terbuka untuk siapa saja. Analisis sentimen adalah teknik yang mengekstrak data opini dan secara otomatis memahami serta memproses data teks untuk mengidentifikasi sentimen pada opini [4].

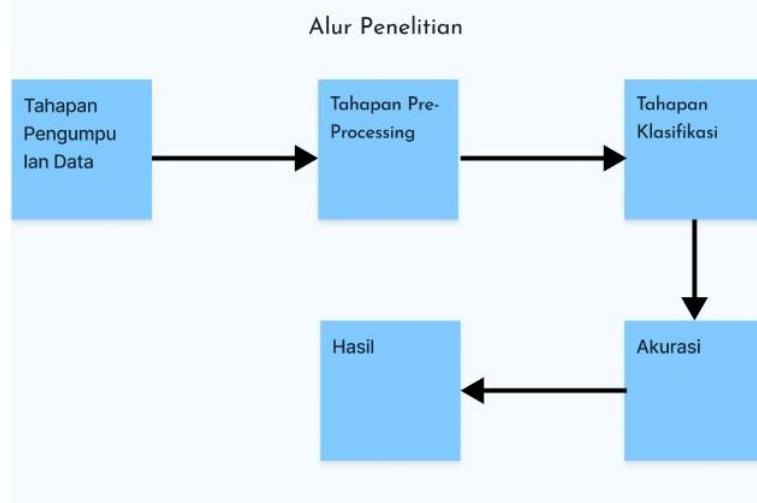
Sentiment analisis digunakan untuk mengetahui sentimen atau polarity sebuah teks apakah *Extremely positive, positive, neutral, negative, Extremely negative*. Biasanya sentiment analysis diterapkan pada data teks opini masyarakat terhadap suatu objek, misalkan review dari suatu produk *e-commerce*, review sebuah film, dan komentar yang terdapat pada sosial media [5]. Masalah utama pada penelitian ini adalah kurangnya pemahaman mengenai faktor-faktor yang mempengaruhi opini publik terhadap kebijakan tarif impor Donald Trump. Meskipun banyak survei dilakukan, data yang dihasilkan sering kali tidak seimbang, dengan dominasi opini tertentu yang dapat menyulitkan analisis prediksi yang akurat. Oleh karena itu, pendekatan statistik dan machine learning diperlukan untuk menggali lebih dalam hubungan antara variabel-variabel demografis, ekonomi, dan politik dengan opini publik. Penelitian ini menarik karena menggabungkan analisis sosial-politik dengan metode kuantitatif, yakni *logistic regression* dan teknik *oversampling SMOTE*, yang jarang diaplikasikan dalam kajian opini publik terkait kebijakan perdagangan. Menurut data dari *Pew Research Center* dan *Gallup*, opini publik Amerika terhadap kebijakan tarif Trump mengalami fluktuasi sepanjang 2018 hingga 2020. Sekitar 45% responden menyatakan mendukung kebijakan tersebut pada tahun 2018, sementara sisanya menolak atau tidak memberikan pendapat. Namun, survei lanjutan menunjukkan peningkatan ketidaksetujuan pada tahun 2019 dan 2020, terutama dari konsumen dan pelaku usaha kecil. Di sisi lain, data demografis menunjukkan bahwa opini publik sangat dipengaruhi oleh faktor seperti afiliasi politik, tingkat pendidikan, dan status ekonomi. Namun, data opini publik ini cenderung tidak seimbang, dengan lebih banyak representasi dari kelompok mayoritas, sehingga dibutuhkan pendekatan khusus untuk mengatasi ketidakseimbangan data tersebut.

Pada penelitian lain yang membahas tentang Analisis Sentimen Masyarakat terhadap Kasus Korupsi PT. Timah yang menggunakan kombinasi metode *Support Vector Machine* dan *SMOTE* menghasilkan akurasi yang cukup tinggi, yaitu sebesar 88% sedangkan tanpa kombinasi *SMOTE* menghasilkan akurasi sebesar 89%. Namun, hanya sedikit penelitian yang menggunakan pendekatan machine learning seperti *logistic regression* dikombinasikan dengan *Synthetic Minority Over-sampling Technique (SMOTE)* untuk mengatasi ketidakseimbangan data dalam analisis opini publik. Penelitian ini mencoba mengisi kesenjangan tersebut dengan menerapkan metode kuantitatif untuk memperoleh hasil yang lebih akurat dan representatif. Kurangnya pemahaman terhadap opini publik yang valid dan representatif dapat menyebabkan kesalahan dalam pengambilan kebijakan di masa depan. Misalnya, jika opini publik yang ditangkap tidak mencerminkan populasi secara keseluruhan, maka kebijakan lanjutan yang berbasis pada data tersebut bisa menimbulkan ketidakpuasan sosial, penurunan kepercayaan publik, dan kerugian ekonomi. Secara teoritis, hal ini juga menghambat pengembangan model prediksi yang dapat diandalkan dalam ilmu sosial-politik.

Peneliti memilih topik ini karena kebijakan tarif impor Donald Trump merupakan contoh konkret dari bagaimana kebijakan ekonomi mempengaruhi dinamika opini publik dalam skala nasional. Selain itu, pendekatan kombinasi antara *logistic regression* dan *SMOTE* masih cukup jarang digunakan dalam kajian semacam ini, sehingga penelitian ini dapat memberikan kontribusi metodologis baru dalam analisis opini publik yang berbasis data tidak seimbang. Metode *SMOTE* adalah terusan dari metode *oversampling*, yang bekerja dengan mengambil sampel baru dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang dengan cara mensampling ulang kelas minoritas [6]. Tujuan dari penelitian ini adalah untuk menganalisis faktor-faktor yang mempengaruhi opini publik terhadap kebijakan tarif impor Donald Trump dengan menggunakan metode regresi logistik dan *SMOTE*. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih dalam mengenai pola opini publik berdasarkan berbagai variabel penentu, serta menyajikan model prediksi yang dapat digunakan dalam studi kebijakan publik dan ilmu politik. Secara praktis, hasil penelitian ini dapat membantu pembuat kebijakan, analis politik, dan peneliti sosial dalam memahami persepsi masyarakat terhadap kebijakan ekonomi yang kontroversial.

2. METODE PENELITIAN

Pada Metode Penelitian ini diawali dengan proses dilakukannya pengumpulan data berupa komentar yang tersebar dari video pada platform Youtube terkait kebijakan tarif Donald Trump. Kemudian, setelah data tersebut menjalani tahap *preprocessing* data, Data diproses untuk menjalankan klasifikasi atau dianalisis opini menjadi positif atau negatif, yang selanjutnya akan dievaluasi menggunakan data uji dengan *matrix precision, recall, F1-Score*, dan akurasi. Selain itu pada tahap klasifikasi akan diberikan juga akurasi dan juga hasil dari proses yang dilakukan. Dapat dilihat pada gambar 1 untuk tahapan atau alur penelitian



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Dalam penelitian ini, pada proses pengumpulan data dilakukan menggunakan data primer yang diperoleh secara langsung dari komentar dari salah satu video pada platform Youtube, yaitu terkait opini masyarakat terhadap kebijakan tarif Donald Trump dengan link : https://youtu.be/Z6T_Arwn15Q?si=HUp3okKzz6dyXh3k. Komentar-komentar ini mencerminkan opini masyarakat secara langsung terhadap kebijakan tersebut dan digunakan sebagai bahan utama dalam proses klasifikasi opini. Data yang telah diambil dan dikumpulkan adalah sebanyak 1466 komentar. Pengumpulan data ini sendiri menggunakan cara *crawling* untuk mengambil sentimen-sentimen yang terdapat pada video youtube tersebut. Hasil pengumpulan beberapa data yang telah dikumpulkan dapat dilihat pada tabel berikut:

Tabel 1. Contoh Sampel Data

Id	Author	Title
1	@pickfallstriani3526	umkm america tersenyum
2	@derizdeprok	gak takut kaya beli aja produk indonesia
3	@vinsmokefadel969	maaf nanya indonesia ekspor aja min
4	@catalogpedia2718	masyarakat as senang krn kebijakan tdk dpt membeli yg perlukan
5	@stevenking2584	iya bg trump mw ngedorong warganya beli produk lokal

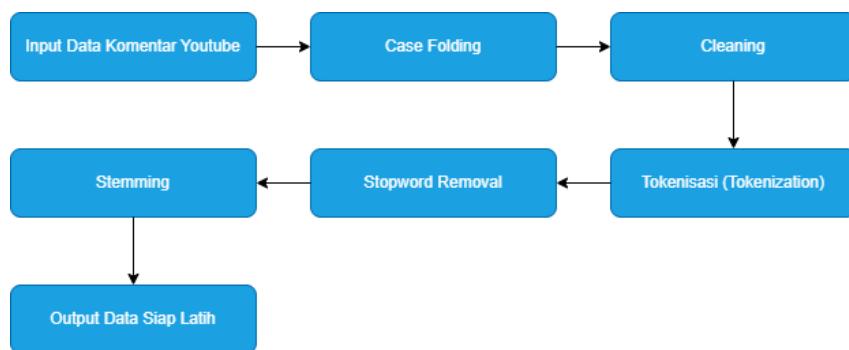
2.2. Text Processing

Text Processing merupakan tahap cukup penting dalam pengolahan data teks sebelum data dilanjutkan untuk digunakan dalam proses pelatihan model klasifikasi. Karena komentar yang diperoleh dari YouTube

berupa data teks mentah, sehingga diperlukan serangkaian langkah pra-pemrosesan untuk membersihkan, menyederhanakan, dan juga mengubah data teks menjadi bentuk yang dapat diolah oleh algoritma *machine learning* sebelumnya. Serangkaian langkah-langkah *pra-processing* yang dilakukan dalam penelitian ini diantaranya yaitu sebagai berikut :

1. *Case Folding*, yaitu pada tahap ini seluruh huruf dalam komentar diubah menjadi huruf kecil (*lowercase*) yang dimana bertujuan untuk menyamakan representasi kata, sehingga kata seperti contoh kata “Trump”, “trump”, dan “TRUMP” diperlakukan sebagai kata yang sama.
2. *Data Cleaning* yang merupakan proses dimana komentar yang dikumpulkan dari video YouTube umumnya mengandung berbagai karakter non-alfabet seperti angka, tanda baca, emoji, URL, simbol, dan mention. Oleh karena itu, maka dilakukan pembersihan teks dengan menghapus unsur-unsur karakter tersebut.
3. *Tokenisasi* atau *Tokenization* adalah proses memecah kalimat menjadi kata-kata individual (token). Setiap komentar diubah menjadi daftar kata untuk memudahkan proses analisis lebih lanjut.
4. *Stopword Removal* adalah kata-kata umum dalam bahasa Inggris seperti “is”, “the”, “and”, “at”, yang tidak memiliki makna penting dalam klasifikasi opini. Kata-kata ini dihapus untuk mengurangi noise dalam data dan memperkuat kata-kata yang bermakna opini.
5. *Stemming* adalah proses mengubah kata menjadi bentuk dasarnya. Misalnya, kata “hurting”, “hurts”, dan “hurted” akan dikembalikan ke bentuk dasar “hurt”. Ini bertujuan untuk menyamakan bentuk variasi kata.

Text Processing ini dilakukan secara konsisten terhadap seluruh data komentar YouTube yang telah diambil dan dikumpulkan yang berupa data yang berkualitas sehingga dapat mempermudah dalam proses klasifikasi. Hasil dari tahap ini kemudian digunakan sebagai input pada model klasifikasi *Logistic Regression*. Alur pra-proses yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Alur Pra-Pemrosesan

2.3. Tahapan Klasifikasi

Klasifikasi data adalah proses menemukan karakteristik yang sama dalam sekumpulan objek dalam database dan mengklasifikasikannya ke dalam kategori yang berbeda menurut model klasifikasi tertentu [7]. Data yang telah diproses sebelumnya akan diklasifikasikan menggunakan metode *Logistic Regression* dan *Smote*. Klasifikasi ini bertujuan untuk mencari keputusan yang terbaik ke dalam sentimen positif dan negatif dari sebuah komentar pada platform YouTube, kemudian setelah itu selanjutnya data akan di *training*. Banyak penelitian telah menunjukkan bahwa *Logistic Regression* adalah metode yang cukup akurat dan telah banyak digunakan untuk klasifikasi teks. Performa model penelitian akan diukur dalam *confusion matrix* seperti pada Gambar 3. *Confusion Matrix*. *Confusion Matrix* adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* Adalah tabel dengan 3 kombinasi berbeda dari nilai prediksi dan nilai aktual. Berikut untuk gambar Confusion Matrix:

		Predicted Values	
		Positive	Negative
Actual Values	Positive	TP	FN
	Negative	FP	TN

Gambar 3. *Confusion Matrix*

Untuk menghitung hasil kinerja model berdasarkan informasi dalam *confusion matrix*, dapat menggunakan dilihat pada persamaan (1), (2), (3) dan (4).

$$precision = \frac{TP}{TP + FP} \dots (1)$$

$$recall = \frac{TP}{TP + FN} \dots (2)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \dots (3)$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \dots (4)$$

Keterangan persamaan (1), (2), (3) dan (4).

TP = True Positif

TN = True Negatif

FP = False Positif

FN = False Negatif

2.4. Logistic Regression

Logistic Regression Adalah jenis regresi analisis yang digunakan untuk menjelaskan hubungan antara variabel dependen dan variabel independen menghubungkan satu atau lebih variabel bebas dengan variabel terikat jenis kategori; bisa 0 dan 1, benar atau salah, besar atau kecil [8]. Berbeda dengan regresi linear yang digunakan untuk prediksi nilai kontinu, *Logistic Regression* digunakan untuk memprediksi probabilitas suatu data masuk ke dalam kelas tertentu, seperti kelas positif atau negatif dan dapat dilihat pada persamaan (5).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \dots (5)$$

Dalam konteks penelitian ini, *Logistic Regression* digunakan untuk mengklasifikasikan opini publik terhadap kebijakan tarif yang dicanangkan oleh Donald Trump, apakah opini tersebut bersifat positif, negatif, atau netral. Model ini mampu menangani variabel input berupa fitur-fitur teks hasil dari proses ekstraksi seperti *TF-IDF* atau *bag-of-words*, dan memberikan output berupa probabilitas terhadap masing-masing kelas.

3. HASIL DAN PEMBAHASAN

3.1. Pre-processing Data

Data preprocessing merupakan hal yang penting namun untuk melakukan data *preprocessing* terhadap teks masih terbilang cukup sulit [9]. Pada tahap ini dilakukan proses *Pre-processing* dari dataset yang diperoleh melalui sentimen-sentimen yang terdapat pada komentar pada platform youtube mengenai kebijakan tarif Donald Trump dengan link berikut : https://youtu.be/Z6T_Arwn15Q?si=HUp3okKzz6dyXh3k, Pada tahap *crawling* didapatkan 1466 komentar data dengan rentang waktu awal mulai video di upload sampai pada 20 Mei 2025. Kemudian dilakukan *text preprocessing* pada dataset tersebut. Seperti pengecilan kapital huruf, penghapusan *stopword*, normalisasi, *tokenisasi*, dan dilakukan *stemming*. Data kemudian disimpan di file terpisah dan hasil *preprocessing* dapat dilihat pada gambar 4.



Gambar 4. Visualisasi hasil data *preprocessing*

seperti yang sudah dilihat pada Gambar 4 yang ada, dataset yang sudah di *pre-processing* akan ditampilkan dalam bentuk *word cloud* yang ditampilkan dalam bentuk token di setiap pengelompokannya yang dimana kata “pajak” muncul di kedua kelompok data kata yang merupakan kata itu sendiri adalah kata pokok atau utama dalam penelitian ini.

3.2. Penerapan Ekstraksi Fitur

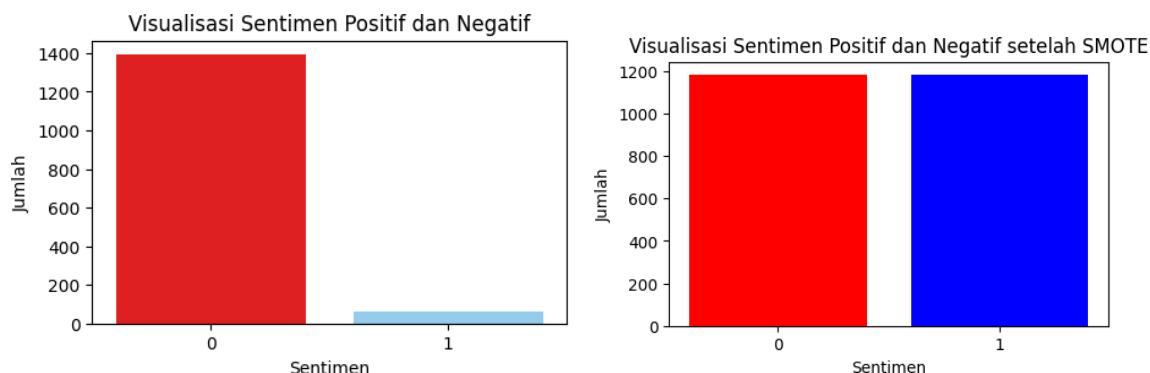
Setelah dilakukannya *preprocessing* data yang dimana kata dari komentar, selanjutnya dilakukannya ekstraksi fitur menggunakan metode *TF-IDF*. *TF-IDF* merupakan suatu metode pembobotan kata yang dikenal baik dalam mengevaluasi pentingnya sebuah kata yang ada dalam dokumen [10]. Dalam menganalisa sentimen, perlu dilakukan ekstraksi fitur fitur yang relevan untuk melihat sentiment analysis yang dipergunakan untuk melihat pendapat atau kesamaan terhadap suatu persoalan atau objek oleh seorang menuju opini yang positif atau negatif. Maka dari itu diterapkannya *TF-IDF* untuk memperhitungkan frekuensi setiap sentimen untuk mendapatkan bobot yang diperlukan untuk tahap klasifikasi. Pembobotan *TF-IDF* ini dilakukan agar pada tahap menentukan klasifikasi sentimen terhadap kebijakan tarif Donald Trump dapat dilakukan dan ditemukan hasilnya. Hasil Ekstraksi Fitur dapat dilihat pada tabel 2.

Tabel 2 Hasil Ekstraksi Fitur

No	amerika	china	indonesia	jadi	malah	mau	negara	nego	sama	tarif
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0.673	0	0	0.739	0	0
2	0	0.000	0	0	0.290	0	0	0.956	0	0
3	0.576	0.736	0.353	0	0	0	0	0	0	0
4	0.576	0.736	0.353	0	0	0	0	0	0	0
...
1464	0	0	0	0	0	0	0	0	0	0
1465	0	0	0	0	0	0	1	0	0	0

3.3. Penerapan Smote

Setelah dilakukannya ekstraksi fitur, dikarenakan terdapatnya ketidakseimbangan dataset yang digunakan, oleh karena itu untuk menyeimbangkan dataset, peneliti menggunakan metode *SMOTE* agar pelabelan antara sentimen positif dan sentimen negatif bisa diseimbangkan. *SMOTE* merupakan metode pembangkitan data minoritas sebanyak data mayoritas [11].



Gambar 5. Menyeimbangkan Dataset

Pembobotan *SMOTE* ini dilakukan agar antara sentimen positif dan sentimen negatif bisa disetarakan. Seperti pada Gambar 5 yang dimana sebelum menggunakan *SMOTE*, perbandingan data antara sentimen positif dan negatif berbeda jauh dan tidak seimbang, setelah dilakukannya *SMOTE*, rasio data mencapai 1:1 antara sentimen negatif dan sentimen positif dimana datanya menjadi seimbang.

3.4. Pengujian dan Hasil Data

Pengujian ini dilakukan dengan menggunakan metode klasifikasi *Logistic Regression*. Dalam penelitian ini seluruh tahapan uji coba menggunakan klasifikasi pada dataset yang telah diolah menggunakan *Python* dengan library *Pandas*.

Tabel 3. Laporan evaluasi dengan *SMOTE*

Akurasi model <i>Logistic Regression</i> : 0.7123287671232876				
	precision	recall	f1-score	support
negatif	0.95	0.74	0.83	210
positif	0.02	0.11	0.03	9
accuracy			0.71	219
macro avg	0.48	0.42	0.43	219
weighted avg	0.91	0.71	0.80	219

Pada tabel 3, terlihat jelas bahwa score yang dimiliki oleh negatif lebih besar dari positif. Misalnya, nilai *Precision* untuk sentimen Negatif mencapai 0.95, sementara untuk Positif hanya 0.02. Begitu juga dengan *Recall*, yang menunjukkan 0.74 untuk Negatif, sementara Positif hanya 0.11. *F1-Score* untuk Negatif juga jauh lebih tinggi, yakni 0.83 dibandingkan dengan 0.03 pada Positif. Jumlah sampel (*support*) pun menunjukkan angka yang jauh lebih besar untuk Negatif dengan 210 sampel, dibandingkan hanya 9 sampel untuk Positif. Selain itu, penelitian yang dilakukan dengan menggunakan *Logistic Regression* dan mendapat akurasi dengan *SMOTE* yaitu 71%.

Tabel 4. Laporan Klasifikasi Tanpa *SMOTE*.

Akurasi model <i>Logistic Regression</i> : 0.9634703196347032				
	precision	recall	f1-score	support
negatif	0.96	1.00	0.98	210
positif	1.00	0.11	0.20	9
accuracy			0.96	219
macro avg	0.98	0.56	0.59	219
weighted avg	0.96	0.96	0.95	219

Pada Tabel 4, dapat dilihat bahwa skor yang dimiliki oleh sentimen Positif lebih tinggi dibandingkan dengan Negatif. Misalnya, nilai *Precision* untuk sentimen Positif mencapai 1.00, sementara untuk Negatif hanya 0.96. Begitu juga dengan *Recall*, yang menunjukkan 1.00 untuk Positif, sementara Negatif hanya 0.11. *F1-Score* untuk Positif juga lebih rendah, yaitu 0.20 dibandingkan dengan 0.98 pada Negatif. Jumlah sampel (*support*) pun menunjukkan angka yang jauh lebih besar untuk Negatif dengan 210 sampel, dibandingkan hanya 9 sampel untuk Positif. Selain itu, penelitian yang dilakukan dengan menggunakan *Logistic Regression* dan mendapat akurasi tanpa *SMOTE* cukup tinggi yaitu 96%.

4. KESIMPULAN

Dari penelitian dan analisis yang telah dilakukan dapat disimpulkan, terlihat bahwa pemakaian *Logistic Regression* yang digabungkan dengan metode *SMOTE* menghasilkan ketepatan atau akurasi sebesar 71% dalam mengklasifikasi bagaimana opini masyarakat tentang aturan tarif impor Donald Trump. Jika dibandingkan, tanpa memakai metode *SMOTE*, tingkat akurasi model mencapai 96%. Hal ini menunjukkan

bahwa meski *SMOTE* berhasil menyeimbangkan data, namun malah membuat performa model yang dilatih jadi kurang akurat. Metode *SMOTE* dipakai untuk mengatasi masalah ketidakseimbangan data antara sentimen positif dan negatif, yang sering menjadi kendala dalam analisis sentimen.

DAFTAR PUSTAKA

- [1] Faiqah, F., Nadjib, M., & Amir, A. S. (2016). "YouTube Sebagai Sarana Komunikasi bagi Komunitas Makassarvidgram", *Jurnal Komunikasi KAREBA*, vol. 5 no. 2, pp. 259-272, 2017, doi : <https://doi.org/10.31947/kjik.v5i2.1905>.
- [2] L. Indrianingsih and Budiarsih, "Analisis Hukum Konten Negatif di Platform YouTube di Indonesia", *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, vol. 2, no. 3, pp. 892–915, 2022, doi: <https://doi.org/10.53363/bureau.v2i3.71>.
- [3] W. A. Maulina, A. Damayanti, "Dampak Tarif Impor Output dan Input Terhadap Probabilitas Perusahaan Keluar Dari Pasar", vol. 13, pp. 47-70, 2018, <https://jurnal.kemendag.go.id/index.php/bilp/article/view/367/221>.
- [4] F. Caroline, R. G. S. Budi, M. E. A. Rivan, "Analisis Sentimen Masyarakat terhadap Kasus Korupsi PT. Timah Menggunakan Metode Support Vector Machine," *Jurnal Ilmu Komputer dan Informatika (JIKI)*, Vol. 4, pp. 44-50, 2024, doi: <https://doi.org/10.54082/jiki.141>.
- [5] Kelvin, J. Banjarnahor, E. Indra, S. H. Sinurat, "Analisis Perbandingan Sentimen Corona Virus Disease-2019 (COVID19) Pada Twitter Menggunakan Metode Logistic Regression dan Support Vector Machine (SVM)", *Jusikom Prima*, vol. 5, no. 2, pp. 47-52, 2022, doi: <https://doi.org/10.34012/jurnalsisteminformasidanimukomputer.v5i2.2365>.
- [6] V. Wijaya dan N. Rachmat, "Penerapan SMOTE dan Regresi Logistik pada Website Skrining Awal Kesehatan Mental Mahasiswa," *Jurnal Algoritma*, vol. 5, no. 2, hlm. 121–130, 2025. doi: <https://doi.org/10.35957/algoritme.v5i2.9046>.
- [7] Ardiansyah, B., Daulay, I., Firdaus, M., Hutagaol, R., & Rahmaddeni. (2023). Analisis Sentimen Opini Publik Terhadap Penerimaan Bantuan Subsidi Upah (BSU) Menggunakan Algoritma Naive Bayes. Prosiding Seminar Nasional Penelitian dan Pengabdian Masyarakat (SENTIMAS), vol. 1, no. 1, pp. 155–162. doi: <https://doi.org/10.58560/sentimas.v1i1.4353>.
- [8] F. D. Pramakrisna, F. D. Adhinata, N. Annisa, and F. Tanjung, "Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression Web-based Classifying SMS Application Using Logistic Regression Algorithm," vol. 11, no. 2, pp. 90–97, 2022, doi: <https://doi.org/10.34148/teknika.v11i2.466>.
- [9] Hakim, B., "Analisa Sentimen Data Text Preprocessing pada Data Mining dengan Menggunakan Machine Learning," *Journal of Business and Audit Information Systems*, vol. 4, no. 2, pp. 16–22, 2021, doi: <http://dx.doi.org/10.30813/jbase.v4i2.3000>.
- [10] Khairunnisa, S., Adiwijaya, dan Al Faraby, S., "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 406–414, 2021, doi: <https://doi.org/10.30865/mib.v5i2.2835>.
- [11] Kasanah, A. N., Muladi, dan Pujiyanto, U., "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 196–201, 2019, doi: <https://doi.org/10.30813/jurnal.iaji.v3i2.196-201>.