

ANALISIS SENTIMEN PUBLIK TERHADAP INVASI ZIONIS KEPADA HAMAS MENGGUNAKAN *SUPPORT VECTOR MACHINE* (SVM)

Dellia Deltania, Garno, Asep Jamaludin

Informatika, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia

2010631170063@student.unsika.ac.id

ABSTRAK

Konflik antara Israel dan Palestina telah menjadi permasalahan yang kompleks dan berlarut-larut selama beberapa dekade, dengan serangan militer Israel terhadap Hamas di Gaza menjadi salah satu momen penting dalam konflik ini. Penelitian ini bertujuan untuk mengidentifikasi dan memahami pandangan masyarakat mengenai konflik Hamas-Israel melalui analisis sentimen menggunakan metode *support vector machine* (SVM) pada data Twitter selama Perang Gaza dan mengembangkan teknik visualisasi yang efektif untuk mempresentasikan sentimen terkait peristiwa yang terjadi selama konflik. Metodologi yang digunakan adalah *knowledge discovery of data* (KDD) yang meliputi *langkah data selection, preprocessing, transformation, data mining, dan evaluation*. Data berupa *tweet* yang diklasifikasikan menjadi positif dan negatif. Hasil penelitian menunjukkan bahwa mayoritas sentimen publik cenderung menentang tindakan kedua belah pihak dalam konflik, dengan adanya ketidakpuasan dan kekhawatiran yang berkelanjutan terhadap situasi tersebut. Penggunaan SVM menghasilkan akurasi tertinggi pada *split data training* 90% dan *testing* 10% sebesar 79.45%, dengan *precision* 63.19%, *recall* 63.63%, dan *F1-score* 63.41%. Algoritma SVM dapat mengklasifikasikan sentimen dengan akurasi yang cukup baik dan visualisasi data memberikan representasi yang jelas tentang distribusi sentimen ini, menegaskan pandangan negatif dalam diskusi publik.

Kata kunci : Analisis Sentimen, Hamas, Support Vector Machine, Visualisasi Data, Zionis.

1. PENDAHULUAN

Konflik antara Israel dan Palestina merupakan salah satu konflik paling kompleks dan lama di dunia, dengan serangan militer Israel terhadap Hamas di Gaza menjadi salah satu peristiwa pentingnya. Hamas, yang didirikan pada akhir 1980-an, adalah organisasi Palestina dengan sayap politik dan militer, dianggap teroris oleh beberapa negara seperti Amerika Serikat dan Uni Eropa, sementara pihak lain melihatnya sebagai gerakan perlawanan sah terhadap pendudukan Israel. Serangan Israel terhadap Hamas sering kali memicu reaksi kuat dari masyarakat internasional, baik dukungan maupun kecaman, tergantung pada pandangan politik, geografis, dan ideologis masing-masing.

Media sosial, terutama Twitter, telah menjadi platform utama bagi publik untuk menyuarakan pandangan mereka tentang konflik ini. Twitter memfasilitasi penyebaran informasi dengan cepat dan luas, memainkan peran penting dalam membentuk opini publik, namun juga membawa tantangan dalam mengelola informasi yang akurat dan seimbang[1][2].

Analisis sentimen menggunakan teknologi *machine learning* telah dilakukan pada berbagai penelitian untuk melihat perbandingan antara sejumlah algoritma populer seperti *Naïve Bayes*, *Support Vector Machine* (SVM), *Decision Trees*, *K-Nearest Neighbor* (KNN), dan *Random Forest*. Dalam konteks analisis sentimen, algoritma SVM sering kali menjadi pilihan utama, memberikan akurasi tinggi dalam mengidentifikasi dan mengklasifikasikan sentimen dari teks. Memberikan gambaran yang lebih

luas mengenai cara teknologi dapat memahami respon manusia terhadap berbagai konteks[3][4][5].

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah proses untuk mengenal perasaan atau emosi seseorang atau sekelompok orang terhadap suatu hal. Hal tersebut dapat berupa produk, layanan, organisasi, individu, isu, kejadian, topik, atau atribut yang terkait dengan suatu fenomena [6]. Fokus pada sentimen positif dan negatif dalam analisis sentimen memiliki beberapa keunggulan utama. Pertama, klasifikasi ini memberikan gambaran awal yang cukup jelas tentang sentimen yang ada dalam teks. Dengan mengenali sentimen positif dan negatif, kita dapat dengan cepat mengidentifikasi pendapat yang mendukung atau menentang suatu topik atau entitas. Hal ini membantu dalam memahami persepsi umum dan sikap masyarakat terhadap produk, layanan, politik, atau topik lainnya. Selain itu, berfokus pada sentimen positif dan negatif memungkinkan pengambilan keputusan yang lebih terinformasi di berbagai bidang, termasuk bisnis, politik, dan keuangan[7].

2.2. Proses pelabelan dan visualisasi data

Proses pelabelan mempermudah model untuk mengategorikan sentimen. Hal ini dikarenakan model membutuhkan input berupa angka dari data teks, sehingga penting untuk memberikan label pada data teks tersebut terlebih dahulu. Dalam konteks analisis sentimen, dengan menentukan tugas dan memasukan sebuah objek secara otomatis mengunduh dan

menggunakan model default untuk menganalisis teks. Hal ini memungkinkan pengguna untuk dengan cepat dan efisien menerapkan analisis sentimen pada berbagai data teks tanpa harus melakukan konfigurasi atau pengaturan tambahan. *Distilbert-base-uncased-emotion* merupakan model yang telah disesuaikan khusus untuk mengidentifikasi berbagai emosi dalam teks, seperti sedih, senang, cinta, marah, takut, dan terkejut yang disediakan oleh *Hugging Face Hub*[8].

2.3. Data Crawling

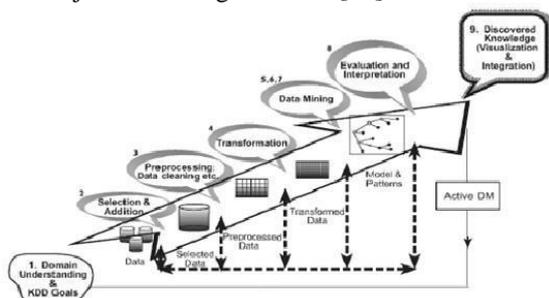
Crawling adalah proses cepat di mana sejumlah besar halaman web disimpan secara lokal dan diindeks berdasarkan sekumpulan kata kunci [9]. Data yang dikumpulkan dalam penelitian ini meliputi data pengguna dan tweet beserta atribut-atributnya yang diunduh dari server Twitter. Pengambilan data pada Twitter melalui proses *crawling* merupakan suatu metode pengunduhan informasi dari server Twitter dengan menggunakan *Application Programming Interface (API)* Twitter, yang dapat berupa data tentang pengguna atau data tweet[10].

2.4. Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) adalah sebuah proses mengubah data teks menjadi data yang numerik dan memberikan bobot pada setiap kata atau fiturnya. TF-IDF adalah metrik statistik untuk mengevaluasi pentingnya kata dalam sebuah dokumen. TF (*Term Frequency*) menunjukkan berapa kali suatu kata tertentu muncul dalam suatu dokumen, sehingga menunjukkan arti kata tersebut dalam berkas tersebut. DF (*Frekuensi Dokumen*) menunjukkan seberapa sering sebuah kata muncul dalam sekumpulan dokumen. IDF (*Inverse Document Frequency*) merupakan kebalikan dari nilai DF. Pada TF-IDF, bobot kata dihitung dengan mengalikan nilai TF dan IDF. Jika suatu kata sering muncul dalam suatu dokumen tertentu, maka kata tersebut memiliki bobot lebih tinggi, namun jika kata tersebut muncul dalam banyak dokumen, maka kata itu memiliki bobot lebih rendah[11].

2.5. Knowledge Discovery of Data (KDD)

Knowledge Discovery of Databases (KDD) adalah metode terstruktur untuk mencari pola yang valid, inovatif, dapat berguna, dan dapat dipahami dalam kumpulan data besar dan kompleks. Proses KDD dijelaskan sebagai berikut [12]:



Gambar 1. Knowledge Discovery of Databases

2.6. Data Selection

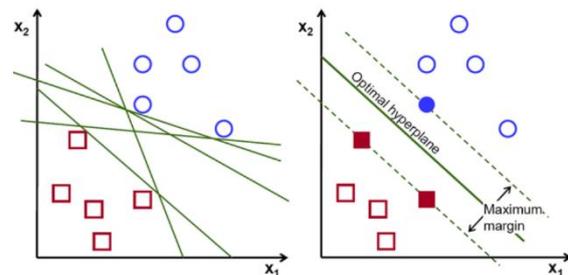
Pemilihan data dari kelompok data operasional harus terjadi sebelum langkah penambangan informasi KDD dimulai[13]. Data terpilih yang digunakan untuk proses data mining disimpan dalam file terpisah dari data operasional.

2.7. Preprocessing

Penting untuk melakukan langkah awal pembersihan data sebelum data mining. Hal inilah merupakan fokus dari *Knowledge Discovery in Databases (KDD)*. Proses pembersihan ini mencakup langkah-langkah seperti menghapus data yang memiliki duplikat, memeriksa keintegritasan data, dan memperbaiki kesalahan yang terjadi pada data seperti *typo*. Selain itu, terdapat proses pengayaan yang bertujuan untuk memperluas informasi pada data yang ada dengan menambahkan data atau informasi tambahan yang diperlukan terkait konteks KDD, seperti data eksternal yang relevan.

Tahap preprocessing dalam penelitian ini terdiri dari beberapa proses, yaitu *cleaning*, *noise removal*, *case folding*, *tokenization* dan *stemming*[14].

2.8. Support Vector Machine (SVM)



Gambar 2. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma yang membangun modelnya dengan mencari hyperplane yang memaksimalkan margin. Hyperplane ini merupakan bidang pemisah dalam ruang berdimensi tinggi yang memisahkan data ke dalam kelas berbeda, dengan support vectors menjadi titik-titik terdekat dari setiap kelas yang dikenali oleh SVM dan berperan penting dalam menentukan hyperplane. SVM memiliki beberapa keunggulan yang signifikan, termasuk kemampuannya dalam penanganan data berdimensi tinggi, efektif mengatasi data dengan banyak fitur, serta kemampuannya dalam mengatasi overfitting dengan parameter yang tepat untuk menghindari overfitting. Selain itu, SVM juga mampu menangani data non-linier melalui konsep transformasi kernel yang memproyeksikan data ke dalam ruang fitur yang lebih tinggi sehingga menjadi linearly separable[15][16].

2.9. Evaluation

Hasil dari proses *data mining* harus disajikan sedemikian rupa sehingga dapat dipahami oleh para pemangku kepentingan. Langkah ini, yang merupakan

bagian dari proses *knowledge discovery of database* (KDD) dan disebut interpretasi, melibatkan pemeriksaan apakah pola atau informasi yang ditemukan sesuai dengan fakta atau hipotesis yang ada[17].

2.10. Confusion Matrix

Selama proses *data mining*, ada beberapa cara untuk mengevaluasi performa model yang dihasilkan, termasuk penggunaan matriks konfusi. Matriks konfusi adalah alat untuk menghitung keakuratan suatu proses penambangan data. Presisi merupakan perbandingan kasus yang nantinya diprediksi positif dengan kasus positif sebenarnya pada data-data yang sebetulnya. Recall atau sensitivitas itu perbandingan antara kasus positif yang teridentifikasi secara benar dengan seluruh kasus positif aktual[18].

Tabel 1 Model Confusion Matrix

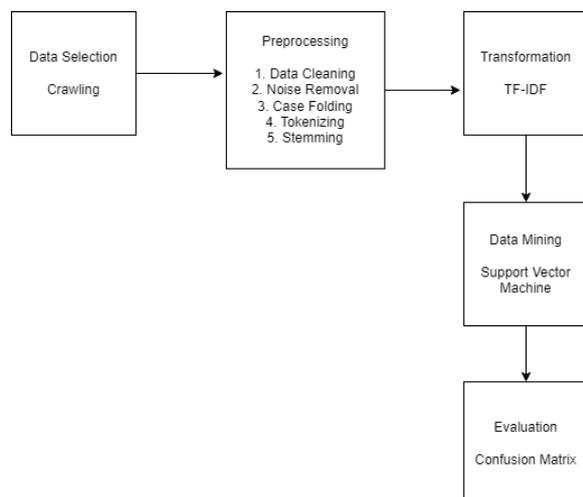
Aktual	Classified as	
	+	-
+	True positive (A)	False negatve (B)
-	False positive (C)	True negative (D)

ROC adalah grafik dua dimensi dengan garis horizontal yang mewakili positif palsu dan garis vertikal yang mewakili positif sebenarnya. AUC (area di bawah kurva) dihitung untuk mengukur perbedaan kinerja metode yang digunakan. Nilai tingkat diagnostik ROC adalah [18]:

1. Akurasi 0,90 – 1,00 = *excellent classification*
2. Akurasi 0,80 – 0,80 = *good classification*
3. Akurasi 0,70 – 0,80 = *fair classification*
4. Akurasi 0,60 – 0,70 = *poor classification*
5. Akurasi 0,50 – 0,60 = *failure*

3. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini yaitu *Knowledge Discovery of Data* (KDD) sesuai dengan gambar 3 dibawah ini:



Gambar 3. Bagan metodologi penelitian

3.1. Data Selection

Pada tahap ini akan melakukan pencarian dan pengambilan data dari media sosial Twitter. Pengambilan data menggunakan teknik crawling data dengan python dari rentang waktu Oktober sampai Desember 2023. Data yang digunakan yaitu tweet berbahasa Indonesia yang berisi opini. Kemudian akan mengambil data dengan memasukan keyword Konflik, Hamas, Israel supaya data yang ada dapat dibagi menjadi dua kelas yaitu sentimen positif dan sentimen negatif.

3.2. Preprocessing

Pada tahap ini akan melakukan Preprocessing berupa data cleansing agar data mentah dan kotor yang diambil melalui Twitter menjadi konsisten dan tidak ada noise. Selain itu, ada pula beberapa tahap preprocessing yang lain, seperti:

- a. *Data Cleaning* akan menghilangkan beberapa data dalam dataset yang memerlukan penanganan melalui berbagai metode seperti menghapus bahasa selain Indonesia, data yang terduplikat, nilai yang hilang, dan penyesuaian tipe data.
- b. *Noise Removal* akan menghilangkan kata-kata penghubung, tanda baca, URL, atau karakter khusus yang tidak diperlukan dalam analisis sentimen. Untuk dokumen berbahasa Indonesia, contoh dari kata-kata penghubung yaitu “di”, “dari”, “yang”, “dan” dan lain – lain.
- c. *Case Folding* untuk mengubah seluruh huruf yang ada dalam dokumen menjadi huruf kecil. Tujuannya agar semua huruf sama dan mudah untuk diolah. Hanya karakter huruf dari 'a' hingga 'z' yang akan diterima. Karakter selain itu akan dibuang.
- d. *Tokenizing* akan membagi teks menjadi token atau kata-kata individual agar sesuai format yang diperlukan oleh model. Untuk memisahkan antar kata akan digunakan spasi.
- e. *Stemming* akan dilakukan untuk mengubah kata-kata dalam sebuah dokumen menjadi bentuk dasarnya dengan cara menghapus semua bagian tambahan seperti awalan, akhiran, sisipan, serta kombinasi dari awalan-akhiran.

3.3. Transformation

Pada tahapan transformasi akan melakukan proses seleksi feature yang memproses perubahan nilai dari kategorical ke numerical. Tahapan tranformasi ini menggunakan TF-IDF supaya proses pembobotan kata mudah dilakukan dimana kata akan di extract kedalam bentuk suatu nilai.

3.4. Data Mining

Pada langkah ini akan melakukan pengolahan data menjadi klasifikasi menggunakan metode algoritma *Support Vector Machine* (SVM). Dalam penelitian ini akan digunakan 2 kelas sentimen akibatnya SVM dengan kernel linear cocok digunakan.

3.5. Evaluation

Tahap evaluation menjadi tahapan terakhir dalam *knowledge discovery of data* (KDD) yang nantinya menghasilkan kesimpulan dari pengolahan data mining. Dalam penelitian ini akan melakukan proses perhitungan akurasi menggunakan *confusion matrix* agar dapat mencari nilai akurasi, *precision*, *recall*, dan *F-measure*.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Tahapan pengumpulan data menggunakan bantuan tools *Google Colab* dan bahasa pemrograman Python. Selain itu, API *tweet-harvest* digunakan untuk membantu proses *crawling*. Data yang sudah terseleksi berjumlah 5103.

4.2. Preprocessing

Tabel 1. Perbandingan sebelum dan sesudah dilakukan preprocessing

Sebelum preprocessing	Sesudah preprocessing
@sumsanghpku @dreimora @tanyakanrl Kalau mau pake logika ya lu bandingin aja siapa paling banyak yg ngerenggut nyawa manusia kalau lu blg hamas kejam oke fair lantas nyawa yg selama ini ditumpahkan oleh Israel g lu jadiin tolak ukur?	pake logika ya lu bandingin aja ngerenggut nyawa manusia lu blg hamas kejam oke fair lantas nyawa tumpah israel g lu jadiin tolak ukur
@deaqct @zqhr Kenapa sih Lo nggak mikir kalo Israel nggak menjajah dulu juga ga bakal ada tuh yang namanya Hamas nyerang? Kalo Lo juga bandingin ratusan atau bahkan ribuan serangan israel sejak 1948 lamanya, serangan Hamas itu ga ada apa2nya coyyy Setdah buzzer apa gimana sih	sih lo nggak mikir kalo israel nggak jajah ga tuh nama hamas nyerang kalo lo bandingin ratus ribu serang israel serang hamas ga nya coyyy setdah buzzer gimana sih
Israel selama ini cuman nunggu serangan hamas terlebih dahulu biar mereka punya alasan jalanin rencana mereka buat pembantaian massal Palestinian krna gk make sense mengatasnamakan self defense" tapi menyerang dan membantai anak fâ€šÅ, Â² tidak bersalah sampai membabi buta begitu."	israel cuman nunggu serang hamas biar alas jalanin rencana bantai massal palestinian krna gk make sense mengatasnamakan self defense serang bantai anak salah babi buta

Tabel 2. Hasil Proses Pelabelan

Kalimat	Label
pake logika ya lu bandingin aja ngerenggut nyawa manusia lu blg hamas kejam oke fair lantas nyawa tumpah israel g lu jadiin tolak ukur	[{'label': 'NEGATIVE', 'score': 0.9929662942886353}]
sih lo nggak mikir kalo israel nggak jajah ga tuh nama hamas nyerang kalo lo bandingin ratus ribu serang israel serang hamas ga nya coyyy setdah buzzer gimana sih	[{'label': 'NEGATIVE', 'score': 0.9988340735435486}]
israel cuman nunggu serang hamas biar alas jalanin rencana bantai massal palestinian krna gk make sense mengatasnamakan self defense serang bantai anak salah babi buta	[{'label': 'NEGATIVE', 'score': 0.9994825124740601}]

Proses ini menghasilkan 3593 data dengan label *NEGATIVE* dan 1509 data dengan label *POSITIVE*. Dengan demikian, tahap ini menandai persiapan data sebelum dilibatkan dalam proses ekstraksi fitur menggunakan metode TF-IDF.

4.3. Transformation

Pada tahap ini, dilakukan ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) setelah data mengalami proses pelabelan menggunakan *Transformers* dari *Hugging Face*.

Gambar 4. Hasil Crawling Data Twitter

Ada 12 atribut yang terdiri dari *created_at*, *id_str*, *full_text*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, dan *tweet_url*. Atribut yang akan digunakan hanya *full_text* yang berisi tweet dari publik.

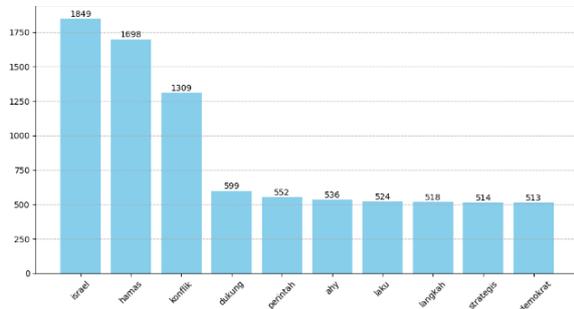
Tabel 3. Jumlah Data yang Dibagi Kedalam *Data Training* dan *Data Testing*

Rasio Perbandingan	Data Training	Data Testing
60:40	3061	2041
70:30	3571	1531
80:20	4081	1021
90:10	4591	511

4.4. Data mining

Pada langkah ini, data yang telah disiapkan dibagi menjadi dua bagian utama, yakni *data training* dan *data testing*. Pembagian ini dilakukan dengan menggunakan fungsi *train_test_split* dengan variasi rasio pembagian, yaitu 60:40, 70:30, 80:20, dan 90:10. Dengan pembagian yang sesuai, diharapkan dapat

Visualisasi *word cloud* ulasan positif. Kata konflik, hamas dan israel mendominasi visualisasi *word cloud* ini. Selain itu, terdapat pula kata seperti dukung, perintah, ahy, dan laku yang juga mendominasi kata pada *word cloud*.



Gambar 8. Grafik Frekuensi Kata Sentimen Positif

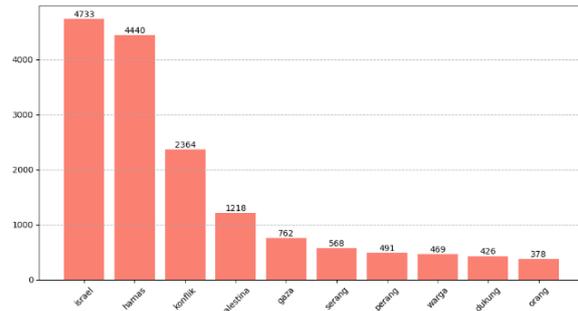
10 kata diatas memiliki frekuensi kemunculan dengan nilai tertinggi dari kata lainnya. Pada diagram ini semakin terlihat kata apa saja yang sering disebutkan pada sentimen positif. Kata israel muncul sebanyak 1849 kali, hamas muncul sebanyak 1698 kali sedangkan konflik muncul sebanyak 1309 kali diikuti dengan kata dukung, perintah, ahy, laku, langkah, strategis, dan pdemokrat secara berurutan. Untuk kata israel dan hamas tidak terlalu jauh frekuensinya pada sentimen positif ini. Pada tanggal 9 Oktober 2023, AHY dari partai Demokrat menyerukan agar konflik tersebut segera dihentikan dan diselesaikan akar permasalahannya. Konflik ini dipandang sebagai isu geopolitik, menunjukkan bahwa fokus utama adalah pada dinamika kekuasaan dan politik. Dalam konteks perang, ada penolakan yang kuat terhadap penggunaan kekerasan dan dukungan terhadap usaha-usaha damai, termasuk melibatkan pihak ketiga yang netral untuk membantu mediasi dan penyelesaian konflik.

Pada ulasan negatif juga terdapat kata yang sering muncul atau tinggi frekuensinya. Berikut merupakan visualisasi *word cloud* dari ulasan dengan sentimen negatif.



Gambar 9. Visualisasi Word Cloud Sentimen Negatif

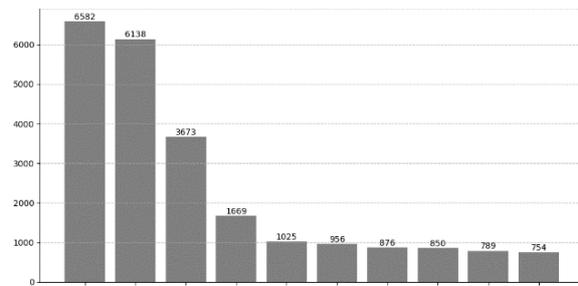
Kata israel, hamas dan konflik juga mendominasi word cloud.



Gambar 10. Grafik Frekuensi Kata Pada Sentimen Negatif

Diagram diatas merupakan diagram frekuensi 10 kata pada sentimen negatif. Kata israel pada sentimen ini muncul sebanyak 4733 kali, hamas sebanyak 4440 kali sedangkan kata konflik menurun drastis sebanyak 2364 kali diikuti dengan palestina, gaza, serang, perang, warga, dukung dan orang secara berurutan. Kata israel juga masih lebih unggul dalam sentimen negatif dibandingkan kata hamas. Dari serangkaian pendapat yang disampaikan, terdapat pandangan konsisten bahwa konflik Hamas-Israel dianggap negatif dan seharusnya dihindari karena membawa kerusakan dan kehancuran. Pendapat juga menunjukkan bahwa keberadaan Hamas dan Israel bisa menimbulkan keresahan di kalangan masyarakat sipil, menandakan penekanan perlunya bantuan untuk damai.

Dari hasil visualisasi *word cloud* tersebut, menunjukkan bahwa ada berbagai kata dengan frekuensi tinggi hingga rendah pada sentimen yang mendominasi. Kata tersebut ditampilkan seperti ini.

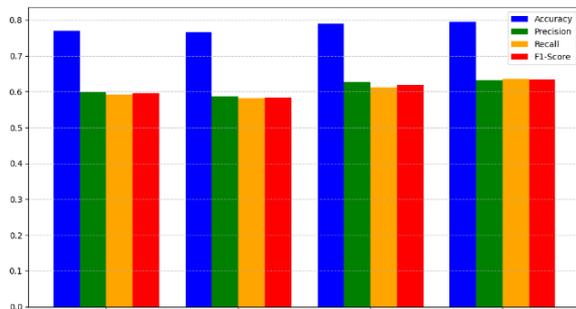


Gambar 11. Grafik Frekuensi 10 Kata Tertinggi

10 kata dengan frekuensi tertinggi yaitu kata “israel” dengan frekuensi sebanyak “6582” kali kemunculan, kata “hamas” sebanyak “6138” kali, kata “konflik” sebanyak “3673” kali, kata “palestina” sebanyak “1669” kali, “dukung” sebanyak “1025” kali, “gaza” sebanyak “956” kali, “perintah” sebanyak “876” kali, “laku” sebanyak “850” kali, “ahy” sebanyak “789” kali, “langkah” sebanyak “754” kali. Secara keseluruhan, terlihat adanya kecenderungan kuat untuk mencari solusi damai dan mengurangi penderitaan yang ditimbulkan oleh konflik ini pada masyarakat umum.

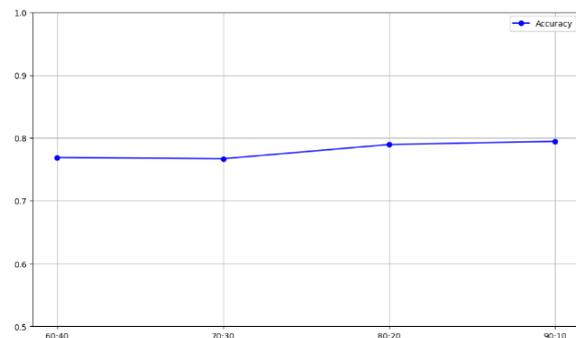
Dalam proses klasifikasi dilakukan 4 skenario pengujian. Skenario pengujian yang dilakukan yaitu

split data training-testing 90% dan 10%, 80% dan 20%, 70% dan 30%, serta 60% dan 40%. Hasil dari skenario pengujian akan ditampilkan pada grafik berikut.



Gambar 12. Grafik Hasil dari Skenario Pengujian

Berikut ini merupakan tampilan dari hasil nilai accuracy yang telah didapatkan.



Gambar 13. Grafik Hasil Akurasi

Setelah dilakukan klasifikasi dengan algoritma SVM menghasilkan akurasi tertinggi di skenario seperti pada gambar yaitu pada pengujian dengan split data training 90% dan data testing 10% sebesar 79.45%. Pengujian skenario terakhir lebih optimal dari skenario lainnya. Pada skenario ini nilai accuracy sebesar 79.45%, precision sebesar 63.19%, recall sebesar 63.63% dan F1-score sebesar 63.41%. Dapat disimpulkan bahwa penelitian dengan nilai akurasi terbaik terhadap Analisis Sentimen Publik Terhadap Invasi Zionis Kepada Hamas Menggunakan Support Vector Machine (SVM) yaitu pengujian split data training 90% dan data testing 10% sebagai model klasifikasi terbaik.

5. KESIMPULAN DAN SARAN

Sentimen publik umumnya menentang tindakan kedua belah pihak dalam konflik Zionis dengan Hamas. Mayoritas sentimen publik menunjukkan ketidakpuasan dan kekhawatiran terhadap konflik ini, dengan kata-kata seperti "Israel", "Hamas", dan "konflik" yang sering muncul. Sentimen publik terhadap konflik ini rata-rata negatif, menunjukkan ketidakpuasan umum masyarakat terhadap tindakan kedua belah pihak dalam konflik berkepanjangan ini. Algoritma Support Vector Machine (SVM) yang digunakan untuk klasifikasi sentimen konflik Hamas-

Israel menghasilkan akurasi cukup baik, dengan akurasi tertinggi 79,45% pada pembagian data latih 90% dan data uji 10%. Masyarakat juga memvalidasi hasil penelitian ini, menunjukkan bahwa model ini dapat menggeneralisasi dari data latih ke kondisi nyata, meskipun masih perlu penyesuaian lebih lanjut.

Visualisasi data seperti grafik dan word cloud memberikan gambaran yang jelas dan mudah dipahami tentang distribusi sentimen, mengonfirmasi dominasi pandangan negatif dalam diskusi publik tentang konflik Hamas-Israel, dan didukung oleh pengamatan di lapangan. Metodologi yang digunakan terbukti mampu menangkap dan memvisualisasikan sentimen terkait konflik dengan akurat. Saran penelitian selanjutnya adalah menggunakan metode lain sebagai perbandingan, seperti Naive Bayes, Decision Trees, atau Random Forests, untuk analisis sentimen yang lebih mendalam mengenai konflik Hamas-Israel menggunakan metode k-means clustering sebelum menerapkan SVM untuk mengelompokkan data terlebih dahulu, sehingga meningkatkan akurasi dan efisiensi klasifikasi, serta menggunakan sumber lain seperti Facebook, Instagram, YouTube, atau media sosial lainnya.

DAFTAR PUSTAKA

- [1] A. T. Chatfield, H. J. J. Scholl, and U. Brajawidagda, "Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services," *Gov. Inf. Q.*, vol. 30, no. 4, pp. 377–386, 2013, doi: 10.1016/j.giq.2013.05.021.
- [2] N. A. Azmi, A. T. Fathani, D. P. Sadayi, I. Fitriani, and M. R. Adiyaksa, "Social Media Network Analysis (SNA): Identifikasi Komunikasi dan Penyebaran Informasi Melalui Media Sosial Twitter," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1422, 2021, doi: 10.30865/mib.v5i4.3257.
- [3] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis," in *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, SciTePress, 2019, pp. 525–531. doi: 10.5220/0008364105250531.
- [4] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 212–218, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [5] R. Yuranda, T. Sutabri, and D. Wahyuningsih, "Machine Learning Approach in Evaluating News Labels Based on Titles: Online Media Case Study," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, pp. 434–439, Nov. 2023, doi: 10.32736/sisfokom.v12i3.1808.
- [6] S. Gusriani, K. D. K. Wardhani, and M. I. Zul,

- “Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka),” *4th Appl. Bus. Eng. Conf.*, vol. 1, no. 1, pp. 1–7, 2016, [Online]. Available: https://www.researchgate.net/profile/Muhammad-Ihsan-Zul/publication/311107797_Analisis_Sentimen_Terhadap_Toko_Online_di_Sosial_Media_Menggunakan_Metode_Klasifikasi_Naive_Bayes_Studi_Kasus_Facebook_Page_BerryBenka/links/583e787b08ae8e63e6158698/Analisis-Se
- [7] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Appl. Sci.*, vol. 13, no. 7, 2023, doi: 10.3390/app13074550.
- [8] F. Pascual, “Getting Started with Sentiment Analysis using Python,” Hugging Face. [Online]. Available: <https://huggingface.co/blog/sentiment-analysis-python>
- [9] F. S. Sholihuda, B. Yuwono, and H. C. Rustamadji, “Pemanfaatan Text Mining Pada Sistem Pengolahan Skripsi Menggunakan Algoritma Naïve Bayes Classifier Dan Simple Additive Weighting,” *Telematika*, vol. 17, no. 2, p. 120, 2020, doi: 10.31315/telematika.v1i1.3379.
- [10] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, “Data Crawling Otomatis pada Twitter,” no. September, pp. 11–16, 2016, doi: 10.21108/indosc.2016.111.
- [11] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [12] W. Gata, “Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS,” vol. 6, pp. 1–13, 2017.
- [13] A. Asroni, H. Fitri, and E. Prasetyo, “Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik),” *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [14] W. Bourequat and H. Mourad, “Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine,” *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 36–44, 2021, doi: 10.25008/ijadis.v2i1.1216.
- [15] R. Khanna and M. Awad, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. 2015. doi: 10.1007/978-1-4302-5990-9.
- [16] J. S.-T. Nello Cristianini, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. doi: <https://doi.org/10.1017/CBO9780511801389>.
- [17] F. A. Rahman, M. I. Desa, A. Wibowo, and N. A. Haris, “Knowledge discovery database (KDD)-data mining application in transportation,” *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 1, no. 978, pp. 116–119, 2014, doi: 10.11591/eecsi.1.357.
- [18] T. Rosandy, “Naive Bayes Vs C4.5 Ke Kelancaran Biaya Tetek Bengkek,” *2016*, vol. 2, no. 01, pp. 52–62, 2016.