

Silvia Damayanti Simamora - Analisis Sentimen Terkait Konflik Palestina – Israel Pada Media Sosial X Menggunakan Algoritma Naïve Bayes Classifier

by Layanan Perpustakaan UHAMKA

Submission date: 22-Jun-2024 03:19PM (UTC+0700)

Submission ID: 2406653190

File name: Silvia_Damayanti_Simamora_TI_Turnitin_3_-_Silvia_Damayanti.docx (795.23K)

Word count: 5119

Character count: 32128



Analisis Sentimen Terkait Konflik Palestina – Israel Pada Media Sosial X Menggunakan Algoritma Naïve Bayes Classifier

Silvia Damayanti Simamora¹, Faldy Irwiensyah^{2*}, Firman Noor Hasan³

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹silviadamayantis01@gmail.com, ²faldy@uhamka.ac.id, ³firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: faldy@uhamka.ac.id

Submitted: 99/99/9999; Accepted: 99/99/9999; Published: 99/99/9999

Abstrak—Konflik yang terjadi antara Palestina-Israel telah berlangsung selama kurang lebih 76 tahun di mana Gerakan zionisme mencoba untuk membangun sebuah tanah air Yahudi di tanah Palestina. Pada bulan Oktober 2023, berita mengenai konflik tersebut kembali muncul ke permukaan dan masih terus berlanjut hingga Juni 2024. Hal ini menjadi sorotan Masyarakat di dunia termasuk Masyarakat yang berada di Indonesia. Pada media sosial X, banyak terdapat komentar serta postingan negatif mau pun positif mengenai konflik antara Palestina-Israel akibat keberlangsungan dari permasalahan yang dialami Palestina hingga saat ini. Dari suatu permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen yang tersebar di media sosial X terhadap konflik antara Palestina-Israel. Data yang diambil dari media sosial X hanya berfokus pada komentar serta postingan yang berada di Indonesia sebanyak 1.715 data. Penelitian ini juga menerapkan algoritma *Naïve Bayes Classifier* dengan perbandingan antara data latih sebanyak 80% dan data uji sebanyak 20% yang sebelumnya melalui proses *pre-processing*. Data yang diperoleh pada penelitian ini memiliki hasil akurasi sebesar 94%, *precision* 91%, *recall* 100%, serta F1-Score sebesar 95%. Analisis yang dihasilkan pada penelitian ini memiliki sentiment yang positif sehingga dapat dikatakan bahwa respon masyarakat Indonesia di media sosial X menunjukkan tanggapan positif yang keberpihakan terhadap Palestina.

Kata Kunci: Analisis Sentimen; Media Sosial X; Naïve Bayes Classifier; Palestina; Israel

Abstract— The conflict between Palestine and Israel has been ongoing for approximately 76 years, during which the Zionist movement has attempted to establish a Jewish homeland in Palestinian territory. In October 2023, news about this conflict resurfaced and has continued up until June 2024. This issue has drawn global attention, including from the Indonesian public. On the social media platform X, numerous comments and posts both negative and positive regarding the Palestine-Israel conflict have appeared as a result of the ongoing challenges faced by Palestine. This study aims to analyze the sentiment expressed on the social media platform X regarding the Palestine-Israel conflict. The data collected focuses solely on comments and posts from Indonesia, totaling 1,715 entries. The study employs the Naïve Bayes Classifier algorithm, with an 80% to 20% ratio of training data to test data, following a pre-processing phase. The results of this study indicate an accuracy of 94%, precision of 91%, recall of 100%, and an F1-Score of 95%. The analysis reveals a positive sentiment, suggesting that the Indonesian public's response on the social media platform X predominantly shows positive support towards Palestine.

Keywords: Sentiment Analysis; Social Media X; Naïve Bayes Classifier; Palestine; Israel

1. PENDAHULUAN

Konflik yang terjadi diantara Palestina - Israel telah berlangsung kurang lebih 76 tahun [1] di mana gerakan atau ideologi zionisme berupaya untuk mendirikan suatu tanah air Yahudi di sebuah tanah Palestina, dimana pada waktu itu wilayah Palestina sudah ditempati oleh mayoritas penduduk Arab, Palestina, Kristen, dan juga kelompok Yahudi [2]. Yahudi zionisme telah merampas hak yang seharusnya dimiliki oleh warga Palestina, dengan mengabaikan sebuah tuntutan mereka, serta menjadikan warga Palestina sebagai sebuah target penyiksaan, penganiyaan, dan juga pembunuhan [3] Pada periode Perang Dunia, imigrasi Yahudi ke Palestina semakin banyak, dan membuat ketegangan pun semakin meningkat karena komunitas Palestina dan Yahudi sudah mulai bersaing untuk menguasai tanah Palestina[4] Pada konflik perebutan wilayah inilah yang memperdalam jurang antara Yahudi dan Palestina, karena konflik ini menjadi terus berlanjut dan pada saat tahun 1960-an muncul sebuah organisasi pembebasan Palestina (PLO) dan Hamas [5].

Hamas sendiri adalah kelompok militan Palestina yang telah menguasai jalur Gaza, menyediakan layanan, membentuk institusi pemerintahan sendiri, kepolisian beserta pengadilan dan sistem kesejahteraan serta menjaga keamanan di wilayah tersebut. Hamas juga telah terlibat dalam konflik dengan Israel dan sudah ditetapkan sebagai sebuah organisasi teroris oleh beberapa negara. Hamas bertanggung jawab atas serangan yang diluncurkan kepada pasukan militer Israel dan telah memperoleh reputasi menjadi sebuah kelompok yang sangat kejam dengan tindakan yang tidak bisa diprediksi. Termasuk terhadap konflik yang terbaru pada bulan Oktober 2023 dan masih terus berlanjut hingga Juni 2024 ini. Konflik ini tidak hanya menyebabkan kekerasan antara dua kelompok, tetapi juga menyebabkan penderitaan bagi masyarakat sipil di kedua sisi konflik Palestina – Israel yang belum selesai sampai saat ini. Hal tersebut membuat banyak masyarakat turun tangan dalam menyebar luaskan informasi lewat media sosial, yang menjadi tempat utama di mana individu dari berbagai latar belakang secara spontan mengekspresikan opini, emosi, dan sikap mereka tentang konflik tersebut [6].

Selain itu, peneliti menggunakan metode analisis sentimen dapat membantu dalam membangun pemahaman yang lebih mendalam tentang dampak psikologis konflik ini terhadap individu dan kelompok dengan mengidentifikasi

dampak emosional yang dialami oleh masyarakat yang terlibat langsung atau tercermin dalam diskusi di media sosial [7].

Analisis sentimen juga bermanfaat dalam diplomasi dan keputusan politik. Oleh karena itu, memasukkan analisis sentimen ke dalam penelitian tentang konflik Palestina-Israel ini memungkinkan peneliti untuk mengambil tindakan yang lebih tepat waktu dan relevan untuk mendukung perdamaian dari konflik tersebut [8].

Tujuan dari penelitian ini untuk mengetahui seberapa banyak tanggapan dari masyarakat di Indonesia terhadap konflik antara Palestina – Israel yang sudah menjadi perbincangan utama di dunia. Kembali memuncaknya konflik tersebut menimbulkan berbagai macam opini masyarakat di dunia. Opini masyarakat tersebut memiliki berbagai macam pandangan, baik dalam pandangan positif maupun negatif terhadap kedua belah pihak. Pandangan masyarakat ini dituangkan dalam berbagai macam platform media sosial yang mereka gunakan, salah satu media yang banyak digunakan ialah media sosial X [9].

X merupakan sebuah platform media sosial yang sebelumnya dikenal sebagai Twitter, merupakan sebuah *website microblogging* atau sebuah pelayanan media sosial yang dikenal dunia dengan memberikan seluruh penggunaannya untuk mengirim sebuah kabar melalui *post*, pesan dengan secara *real-time*, mengirim gambar atau video, yang membuat seluruh pengguna mampu saling berkomunikasi satu dengan lainnya. Pada platform X ini yang akan digunakan untuk mengambil data *post (tweet)* tentang Konflik antara Palestina-Israel dan akan diolah maupun dianalisis agar mengetahui seberapa banyak pandangan dan sentimen masyarakat di Indonesia yang menggunakan X pada konflik Palestina-Israel dengan menentukan *Naïve Bayes* sebagai algoritma yang akan dipakai untuk mengklasifikasikan sentimen terhadap konflik tersebut [10].

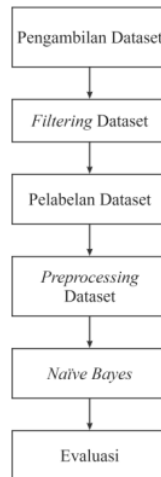
Terdapat penelitian sebelumnya yang menggunakan algoritma *Naïve Bayes* ini untuk menganalisis sebuah sentimen. Peneliti Ade Tiara Susilawati, Nur Anjeni Lestari dan Puput Alpria Nina yang menerapkan algoritma *Naïve Bayes* untuk menganalisis boikot produk israel di Twitter dengan menggunakan dataset sebesar 303 data dengan menghasilkan nilai akurasi sebesar 95%, *precision* sebesar 96%, *recall* sebesar 95% dan F1-Score sebesar 95% [11]. Selanjutnya peneliti Muhammad Yasir, Marissa Grace Haque, Robertus Suraji, dan Istianingsih yang menerapkan algoritma *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (KNN)* untuk menganalisis Konflik antara Palestina dan Israel tentang pemboikotan produk yang terafiliasi Israel menghasilkan sebuah akurasi *Naïve Bayes* 75%, *Decision Tree* 65%, *Random Forest* 67%, SVM 63%, K-NN 53% [12]. Selanjutnya peneliti Rizki Azhar dan Muhammad Farid Wijayanto menerapkan algoritma SVM dengan topik mengungkapkan persepsi dan emosi publik terhadap Konflik Palestina-Israel dengan hasil akurasi sebesar 88%, *precision* sebesar 89% dan *recall* sebesar 90% [13]. Selanjutnya penelitian oleh Hafiz Irsyad dan Akhsani Taqwiym dengan topik tentang Rakyat Palestina dengan menghasilkan akurasi 75% [14]. Peneliti terakhir adalah Fadhillah Meisya Carina, Admi Salma, Dony Permana, dan Zamahsary Martha dengan menerapkan algoritma SVM terhadap Konflik Palestina dan Israel dengan hasil akurasi sebesar 93%, *precision* sebesar 92,93 dan *recall* sebanyak 100%. Dapat dilihat peneliti sebelumnya menunjukkan sebuah persamaan pada saat media sosial twitter yang diambil datanya menggunakan *tweet-harvest* sebagai tempat untuk menganalisis dan mengambil data. Kelima peneliti tersebut sama-sama membahas topik terkait dengan Palestina dan Israel.

Penelitian ini akan kembali melakukan pengolahan data terhadap topik analisis sentimen dari dataset terbaru terkait konflik Palestina dan Israel, yang membandingkan penelitian terdahulu dengan penelitian ini adalah metode yang dipakai hanya berfokuskan pada satu metode yaitu *Naïve Bayes Classifier*, dengan tujuan agar mengetahui apakah metode *Naïve Bayes* memberikan hasil yang lebih baik dan akurat dibandingkan dengan metode lainnya dan pada beberapa penelitian sebelumnya tools yang dipakai untuk mengolah sebuah datanya menggunakan *tools orange*, sedangkan pada penelitian ini *tools* yang digunakan yaitu *RapidMiner*. Penelitian ini juga diharapkan dapat menambah literatur tentang analisis sentimen yang terkait dengan konflik internasional, dengan penekanan khusus pada konflik Palestina-Israel. Selain itu, diharapkan dapat memberikan pemahaman yang lebih baik tentang bagaimana analisis sentimen dapat digunakan secara efektif untuk memahami dinamika opini publik dalam konteks yang berpotensi mempengaruhi kebijakan yang terjadi pada saat ini [15].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode algoritma yang diterapkan dalam penelitian ini adalah menggunakan *Naïve Bayes Classifier*. Terdapat beberapa proses analisis sentimen dengan mengumpulkan sebuah dataset dengan media sosial X terkait dengan topik yang berkaitan dengan Konflik Palestina – Israel. Dataset tersebut kemudian akan diolah dan hasilnya akan dapat diketahui bagaimana tanggapan atau sentimen dari pengguna media sosial X apakah bernilai positif dan negatif [16].



Gambar 1. Flowchart Alur Penelitian

Penelitian ini dilakukan dengan beberapa tahapan yang dapat dilihat pada gambar 1 diatas sebagai berikut:

2.1.1 Pengumpulan Dataset

Pada penelitian ini, dataset diambil dari media sosial X, yang diambil dengan memanfaatkan penggunaan alat pengumpulan *tweet-harvest* untuk menarik data menggunakan platform *Google Colaboratory*. Pada proses pengambilan data ini dilakukan selama 7 bulan secara berkala dan data yang terkumpul sebanyak 4.400 data dan disimpan dalam format data CSV dapat dilihat di gambar 2.



Gambar 2. Proses Ekstraksi Dataset

Google Colab digunakan untuk mengambil sebuah data dari media sosial X dengan cara mengautentikasikan akses ke media sosial X dan memberikan kode_auth dari twitter tersebut untuk menghubungkan antara *Google Colab* dan Media Sosial X dan setelah terhubung dapat dilanjutkan dengan mengambil data *post (tweet)* dengan keyword yang ditentukan. Setelah mendapatkan nantinya akan tersimpan dataset dalam bentuk CSV yang dapat dilihat di gambar 2, setelah dataset sudah disimpan kedalam bentuk CSV [17].

2.1.2 Filtering Dataset

Dataset yang sudah dikumpulkan dengan mengaplikasikan *tweet-harvest* dari *google colaboratory* ditemukan banyak data yang sangat tidak berkaitan dengan topik yang di inginkan dan juga terdapat banyak bahasa asing yang akhirnya diperlukan proses *filtering* ini agar data dapat sesuai dengan standar atau topik pada penelitian ini [18].

2.1.3 Preprocessing Data

Preprocessing data merupakan tahapan awal dari analisis sentimen, pada proses *preprocessing* data terjadi pemrosesan data sehingga siap dianalisis, dalam proses ini setiap kata akan diolah kembali ke dalam beberapa tahap supaya memiliki arti yang lebih sempit. Pada Tahapan *preprocessing* data terdiri dari *replace (remove) mention*, simbol, URL, trim, *remove duplicate*, filter examples, nominal to text, case folding, tokenization, filter stopword, dan filter token by length. Pada tahap ini digunakan tools *RapidMiner* seperti pada Gambar 3.



Gambar 3. Tools RapidMiner

RapidMiner adalah perangkat lunak yang mendukung cara yang lebih mudah untuk memuat data yang dilengkapi dengan Graphical User Interface (GUI) Window untuk memilih set data. Tools RapidMiner seperti Digambar 3 digunakan untuk memeriksa data secara kuantitatif dan kualitatif untuk mendapatkan informasi dan pengetahuan yang diharapkan [19].

2.1.4 Pelabelan Dataset

Setelah tahap pelabelan, data dapat dianalisa dengan cara memberikan label pada setiap ulasan komentarnya. Pada penelitian ini menggunakan dua pelabelan, yaitu *positive* dan *negative*. Pada setiap pelabelan ini sangat penting untuk menentukan akurasi pada setiap prosesnya [19].

16

2.2 Naïve Bayes Classifier

Naïve Bayes merupakan salah satu algoritma pada *machine learning*, yang merupakan sebuah metode untuk mengklasifikasikan yang mudah dengan di dasarkan pada probabilitas sederhana yang di sederhanakan pada teorema Bayes dengan memprediksi sebuah peluang pada masa yang akan datang berdasarkan sebuah pengalaman sebelumnya [20]. Metode Naïve Bayes Classifier mempunyai ciri khas utama yaitu asumsi yang sangat kuat (naif) tentang independensi dari suatu kondisi atau kejadian. Namun, asumsi tersebut bisa jadi tidak benar, tetapi bisa jadi secara praktis algoritmanya membuahkan hasil yang sangat baik [21].

Rumus umum Teorema Bayes dasar dari formula umum Naïve Bayes dapat dilihat sebagai berikut:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \quad (1)$$

Dengan $P(H|E)$ merupakan sebuah probabilitas dari hipotesis H jika *evidence* E terjadi, sedangkan $P(E|H)$ merupakan sebuah probabilitas yang munculnya *evidence* E, ketika hipotesis H terjadi, lalu untuk $P(H)$ merupakan sebuah probabilitas sebuah hipotesis H tanpa mengandung apapun *evidence*, dan untuk $P(E)$ merupakan sebuah probabilitas dari *evidence* E tanpa memandang apapun [22].

2.3 Confusion Matrix

Confusion matrix adalah sebutan dari fundamental pada *machine learning*. *Confusion matrix* telah diterapkan dalam *machine learning* untuk menetapkan ukuran pada akurasi nilai model dengan perbandingan nilai prediksi dengan nilai aktual [23]. Pengukuran model pada penelitian ini menggunakan *matrix* seperti akurasi *Recall*, *Precision* dan *F1-Score*. *Confusion matrix* adalah sebuah tabel yang dapat memprediksi suatu performa algoritma [24]. Penerapan pada tabel *confusion matrix* dapat dilihat seperti pada tabel 1 berikut:

Tabel 1. Confusion Matrix

No.		Prediksi Positif	Prediksi Negatif
1.	Aktual Positif	False Positif	False Negatif
2.	Aktual Negatif	True Positif	True Negatif

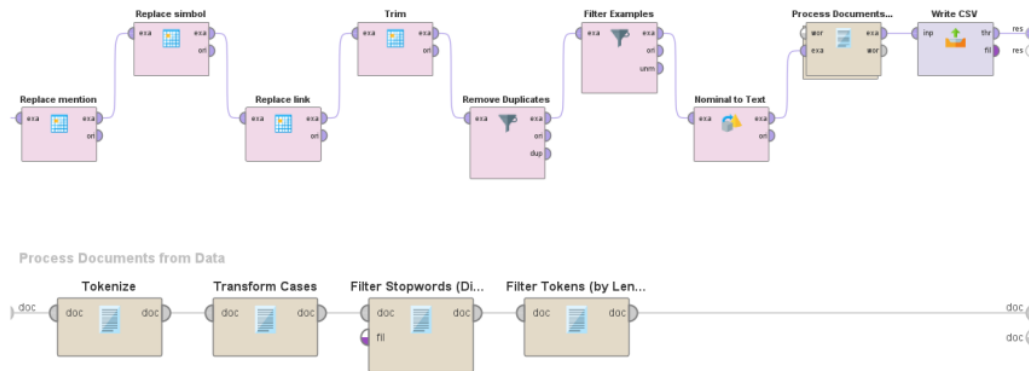
Dalam konteks analisis model prediktif, ada beberapa metrik penting yang perlu dipahami. Pertama, "TP" adalah kumpulan data dari kelas positif aktual yang dihasilkan sebagai positif oleh model. Hal ini menunjukkan seberapa sering model membuat prediksi yang benar padahal data sebenarnya berasal dari kelas positif. "TN" adalah jumlah data dari kelas negatif aktual yang dihasilkan sebagai negatif oleh model. Nilai ini menunjukkan seberapa baik model

12

Setelah sudah terkumpul dataset seperti gambar 3 dan 4, kemudian dataset disatukan pada satu file dengan format CSV dan dilakukan proses filtering dataset untuk menghilangkan bahasa asing dikarenakan pada penelitian ini hanya difokuskan untuk bahasa Indonesia saja.

3.2 Pre-Processing Data

Tahapan selanjutnya adalah preprocessing dataset dengan menggunakan software RapidMiner. Langkah ini sangat penting agar dapat mempersiapkan sebuah data teks yang sudah dimiliki agar lebih optimal untuk dilakukan pada proses selanjutnya. Pada tahap ini bisa dilihat pada gambar 6.



Gambar 6. Proses Pre-Processing Dataset

Ada beberapa tahapan *preprocessing* yang terlihat pada gambar 6 diatas sebagai berikut:

3.2.1 Replace Mention

Dilakukan agar username atau nama pengguna yang masih tersimpan di dataset bisa dihilangkan dengan menggunakan "@" pada *replace* parameter.

Tabel 2. Hasil Replace Mention

Sebelum	Sesudah
@mgicallmemgi @agnezmo Israel palestina bukan konflik agama kocak stop mikir gini.	Israel palestina bukan konflik agama kocak stop mikir gini.
@syauqi_hafiz Mas @syauqi_hafiz tau kan ya stance saya soal konflik israel-Palestina.. Itu belum berubah inshaAllah. Tapi yang ICC lakukan sudah oke (meski tidak ideal). Kapan2 kita diskusi ya :)	Mas tau kan ya stance saya soal konflik israel-Palestina.. Itu belum berubah inshaAllah. Tapi yang ICC lakukan sudah oke (meski tidak ideal). Kapan2 kita diskusi ya :)
@OskarSc80 @HH__ONE Enteng bgt klo nuduh ya Apakah saya setuju dg genosida Israel di Palestina? Enggak tp konteks yg dibahas dri tadi itu konflik SCS kesampingkan konflik Israel-Palestina dulu poin yg kita perdebatkan kali ini ya konflik SCS	Enteng bgt klo nuduh ya Apakah saya setuju dg genosida Israel di Palestina? Enggak tp konteks yg dibahas dri tadi itu konflik SCS kesampingkan konflik Israel-Palestina dulu poin yg kita perdebatkan kali ini ya konflik SCS

Pada tahap ini dapat dilihat di tabel 2 bahwa pada dataset sebelumnya masih banyak teks yang masih memiliki beberapa atribut tidak penting, yaitu *mention* atau *username* dan ketika dilakukan tahapan *replace mention* terlihat perubahan teks menjadi lebih jelas karena sudah tidak ada atribut seperti mention dan username.

3.2.2 Replace Simbol

Dalam proses ini dilakukan agar kata-kata yang tidak penting seperti "_", "!", "~", "?", ":", "@", "(", ")", "µ" dan masih banyak lagi simbol yang ada pada dataset bisa dihilangkan agar data menjadi optimal.

Tabel 3. Hasil Replace Simbol



Sebelum	Sesudah
+ Sebenarnya KONFLIK nih tuh ya KONFLIK PEREBUTAN WILAYAH - Sy jujur aja saya mah ga terlalu mantengin soal konflik Israel vs Palestina ini Bio lu ganti: Sedikit bicara banyak mengetik i. • SEDIKIT MEMBACA MAKANYA GOBLOK i Bismillah ðŸ†µðŸ†µ, Karena sebentar lagi Idul Adha dan genosida Palestina terus berlanjut. Aku berinisiatif bikin campaign Shadaqah Daging untuk Palestina yang tujuannya agar saudara-saudara kita bisa sama-sama makan daging di momen Idul Adha. https://t.co/11F2j651ha Izin #zonauang	Sebenarnya KONFLIK nih tuh ya KONFLIK PEREBUTAN WILAYAH Sy jujur aja saya mah ga terlalu mantengin soal konflik Israel vs Palestina ini Bio lu ganti Sedikit bicara banyak mengetik SEDIKIT MEMBACA MAKANYA GOBLOK Bismillah Karena sebentar lagi Idul Adha dan genosida Palestina terus berlanjut Aku berinisiatif bikin campaign Shadaqah Daging untuk Palestina yang tujuannya agar saudara-saudara kita bisa sama-sama makan daging di momen Idul Adha

Pada tahap ini dapat dilihat di tabel 3 bahwa pada dataset sebelumnya masih banyak teks yang masih memiliki beberapa atribut simbol yang tidak penting, yang akan mempengaruhi hasil data nantinya, sehingga perlu dilakukan tahap *replace simbol* yang berfungsi untuk membersihkan simbol-simbol yang tidak penting.

3.2.3 Replace URL

Digunakan untuk menghilangkan link yang bisa menyebabkan gangguan pada data yang akan diuji nantinya.

Tabel 4. Hasil *Replace URL*

Sebelum	Sesudah
Genosida Gaza Makin Ngeri Total Warga Palestina Mati Tembus 35.000 https://t.co/ShyCBviZ4g	Genosida Gaza Makin Ngeri Total Warga Palestina Mati Tembus 35.000
Konflik Israel vs Pakestina dimulai sejak negara Israel menduduki wilayah Palestina https://t.co/RAI448rJFj	Konflik Israel vs Pakestina dimulai sejak negara Israel menduduki wilayah Palestina

Pada tahap ini dapat dilihat di tabel 4 bahwa pada dataset sebelumnya masih banyak teks yang masih terdapat *link URL* yang akan mempengaruhi hasil data nantinya, sehingga perlu dilakukan tahap *replace URL* yang berfungsi untuk membersihkan link agar data lebih optimal.

3.2.4 Trim

Dilakukan untuk menghapus sebuah spasi tambahan yang berada pada awal dan akhir teks dengan memiliki tujuan untuk menjaga teks data menjadi bersih dan rapih.

3.2.5 Remove Duplicates

Digunakan untuk menghapus data teks yang sama atau double pada dataset. Dilakukan agar setiap data tidak memiliki kesamaan untuk mendapatkan sebuah hasil analisis yang tidak bias dan akurat.

3.2.6 Filter Examples

Dilakukan untuk memfilter data yang tidak relevan pada topik yang diinginkan yaitu Konflik Palestina-Israel.

3.2.7 Nominal to Text

Dapat digunakan untuk memproses sebuah atribut nominal menjadi teks digunakan agar pada proses dokumen *from* data dapat tetap berjalan.

3.2.8 Tokenize

Operator ini digunakan untuk memecah sebuah teks menjadi sebuah token dengan sebuah kata yang bersifat individual seperti kata "palestina", "israel", "aamiin", "anak", "bantuan" dan banyak lagi.

Tabel 5. Hasil Operator Tokenize

Sebelum	Sesudah
Lo pikir orang palestina yang kena genosida itu nyebar berita lewat mana selain Twitter Press Pada dibunuh jurnalisnya TV Bias ke israel Youtube Tiktok IG Pada kena banned Lo mau hapus ya hapus aja gak usah ngajak yang masih mau nyebarin berita kekejaman israel	'Lo', 'pikir', 'orang', 'palestina', 'yang', 'kena', 'genosida', 'itu', 'nyebar', 'berita', 'lewat', 'mana', 'selain', 'Twitter', 'Press', 'Pada', 'dibunuh', 'jurnalisnya', 'TV', 'Bias', 'ke', 'israel', 'Youtube', 'Tiktok', 'IG', 'Pada', 'kena', 'banned', 'Lo', 'mau', 'hapus', 'ya', 'hapus', 'saja', 'gak', 'usah', 'ngajak', 'yang', 'masih', 'mau', 'nyebarin', 'berita', 'kekejaman', 'israel'

Pada tahap ini dapat dilihat di tabel 5 yang menunjukkan bagaimana teks menjadi lebih mudah untuk menganalisis setiap unit kata. Setiap kata dalam teks asli berfungsi sebagai entitas unik yang dapat digunakan untuk berbagai tujuan, seperti analisis frekuensi kata, analisis perasaan, dan sebagainya. Tokenisasi ini membantu dalam mengidentifikasi dan memahami lebih dalam setiap aspek teks yang kompleks.

3.2.9 Transform Cases

Operator ini digunakan untuk mengubah semua huruf yang berada di dalam dataset menjadi huruf kecil semua.

Tabel 6. Hasil Operator Transform Cases

Sebelum	Sesudah
Lo pikir orang palestina yang kena genosida itu nyebar berita lewat mana selain Twitter Press Pada dibunuh jurnalisnya TV Bias ke israel Youtube Tiktok IG Pada kena banned Lo mau hapus ya hapus aja gak usah ngajak yang masih mau nyebarin berita kekejaman israel	lo pikir orang palestina yang kena genosida itu nyebar berita lewat mana selain twitter press pada dibunuh jurnalisnya tv bias ke israel youtube tiktok ig pada kena banned lo mau hapus ya hapus aja gak usah ngajak yang masih mau nyebarin berita kekejaman israel

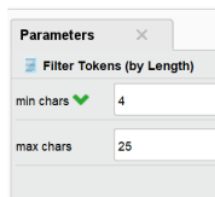
Pada tabel 6 menunjukkan hasil transformasi ini, yang mengubah semua huruf besar menjadi huruf kecil. Hal ini penting untuk memastikan bahwa teks dapat dianalisis secara konsisten tanpa perbedaan yang disebabkan oleh kasus huruf, yang meningkatkan akurasi dan konsistensi dalam proses analisis data berikutnya.

3.2.10 Filter Stopwords

Operator ini digunakan untuk menghapus sebuah kata yang bersambung agar bisa terfokuskan dengan kata yang memiliki makna atau informasi signifikan di dalam sebuah teks yang sedang dilakukan pada sebuah analisis. Pada proses ini diambil data stopwords bahasa Indonesia pada format csv pada platform kaggle.com [26].

3.2.11 Filter Tokens (By Length)

Pada operator ini digunakan untuk menghapus kata yang pendek atau terlalu panjang pada dataset, dengan menggunakan minimal 4 huruf dan maksimal 25 huruf saja seperti pada gambar 7 berikut.



Gambar 7. Parameters Filter Tokens

Hal tersebut menyebabkan kata-kata yang kurang dari 4 dan lebih dari 25 huruf akan dihapus. Setelah melakukan *Pre-processing* Dataset, data yang awalnya sebanyak 4.400 di dalamnya sekarang hanya tinggal 1.715 data saja dan pada saat diproses *remove duplicates*, data yang tersisa hanya sebanyak 1.393 data saja.

3.3 Pelabelan Dataset

Tahapan yang dilakukan selanjutnya adalah memberikan label sentimen dengan menggolongkan dua kategori yaitu sentimen *positive* dan *negative* pada dataset dengan jumlah data yang akan diisi secara manual sebesar 1.393 dan pelabelan manual tersebut dilakukan oleh peneliti dan 3 orang mahasiswa/i dari jurusan Teknik Informatika dan 1 mahasiswa jurusan Bahasa Indonesia dari Universitas Muhammadiyah Prof. Dr. Hamka. Hasil Klasifikasi dapat ditunjukkan pada tabel.7



Tabel 7. Contoh Hasil Pelabelan Sentimen Pada Dataset

Opini	Sentimen
pikir orang palestina yang kena genosida nyebar berita lewat mana selain twitter press pada dibunuh jurnalisnya bias israel youtube tiktok pada kena banned hapus hapus usah ngajak yang masih nyebarin berita kekejaman israel.	Positive
palestina hebat banget bertahan dalam situasi konflik dimata dunia indonesia tidak kalah hebat merdeka mempertahankan tanah tumpah darah.	Positive
kalo orang pendukung palestina tidak boleh main twitter medsosnya meta berarti orang orang gaak boleh tau perkembangan genosida Palestina	Positive
guys kepikiran kemarin habis baca komen orang luar yang bilang kalo Indonesia harusnya malu karena bela belain palestina karena genosida tapi sendirinya melakukan genosida dipapua barat aku awam banget soal begini emang bener	Positive
bayangkan dirimu tinggal hidup gaza kondisi diblokade musuh	Positive
kirinkan gaza keneraka	Negative
saya mendukung israel menghancurkan gaza mengakhiri hamas	Negative
cukup dengan ketidak pedulian sendiri dann sekarang tambah dengan nyinyir ke orang yang lagi berjuang kalo gitu bukan nggak peduli lagi tapi udah pro genosida boikot berhenti membantu zionis terangterangan lewat base ini makan minum darah anak palestina dengan bangga	Negative

Tabel.7 memperlihatkan bahwa pada konflik Palestina-Israel menunjukkan bahwa ada lebih banyak tanggapan positif terhadap pendukung Palestina dapat dilihat pada tabel 8 dan mengkritik atas Tindakan yang dilakukan oleh Israel. Dukungan yang kuat ini mencerminkan empati yang mendalam dan solidaritas internasional terhadap penderitaan yang dialami oleh rakyat Palestina.

Tabel 8. Hasil Klasifikasi Algoritma Naïve Bayes

Identitas	Kategori		
	Positive	Negative	Total
Konflik Palestina - Israel	889	503	1393

Pada tabel 8 menunjukkan hasil klasifikasi menggunakan algoritma Naïve Bayes untuk analisis sentimen terhadap konflik Palestina-Israel. Dari 1.393 tweet yang dianalisis, 889 dianggap positif, dan 503 dianggap negatif. Ini menunjukkan bahwa sekitar 63.84% dari tweet menunjukkan sentimen positif terhadap konflik Palestina-Israel, sedangkan sekitar 36.16% menunjukkan sentimen negatif. Data yang dikumpulkan dari platform media sosial X digunakan dalam analisis ini untuk memberikan wawasan penting tentang bagaimana masyarakat melihat konflik tersebut.

Setelah melakukan pelabelan, dilakukan proses dengan menggunakan metode ²⁴ Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengetahui frekuensi nilai kata atau bobot kata tertentu secara signifikan. Jika suatu kata sering muncul maka semakin tinggi nilai yang dihasilkan [27].

Tabel 9. Hasil TF-IDF

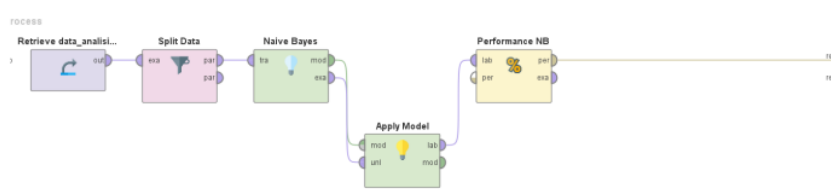
No	Terms				
	adab	adegan	admin	...	akhlak
1.	0	0	0	...	0

2.	0	0	0	...	0
3.	0	0	0	...	0
4.	0	0	0	...	0
5.	0	0.286	0	...	0
6.	0.302	0.281	0	...	0
7.	0	0	0.662	...	0
...
1.709	0	0	0	...	0

Hasil perhitungan TF-IDF untuk masing-masing term atau kata dalam kumpulan ditunjukkan dalam Tabel 9. Misalnya, term-term seperti "adab", "adegan", "administrasi", dan "akhlak" dipelajari untuk mengetahui seberapa sering dan penting mereka muncul di berbagai dokumen. Pada beberapa baris pertama dokumen, kata-kata tersebut memiliki nilai 0, yang menunjukkan bahwa mereka tidak muncul di dalamnya. Namun, pada dokumen kelima, kata "adegan" memiliki nilai TF-IDF 0.286 dan 0.281, masing-masing menunjukkan bahwa kata tersebut muncul beberapa kali dalam dokumen tersebut dengan tingkat signifikansi tertentu.

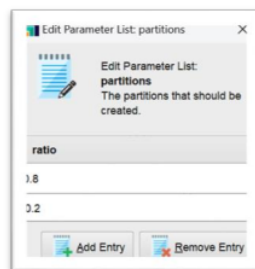
3.4 Implementasi Algoritma Naïve Bayes Classifier

Pada tahapan ini digunakan algoritma Naïve Bayes Classifier untuk menguji data, dimana pada data pelatihan ini dapat membuat algoritma diajarkan dengan menggunakan data latih untuk menemukan model yang sesuai karakteristik dataset, sedangkan data uji dapat mengevaluasi kinerja model yang telah melalui proses pelatihan [28].



Gambar 8. Implementasi Algoritma Naïve Bayes Classifier

Pada titik ini, data diuji dengan algoritma Naive Bayes Classifier, seperti yang ditunjukkan dalam Gambar 8, proses dimulai dengan node "Retrieve data_analisis" mengambil data analisis, kemudian untuk data ini dibagi menjadi dua bagian, data latih dan data uji, melalui proses pemisahan yang dilakukan oleh node "Split Data". Data latih digunakan untuk melatih algoritma Naïve Bayes, yang membuat model yang sesuai dengan karakteristik dataset. Setelah proses pelatihan selesai, node "Apply Model" digunakan untuk mengaplikasikan model yang dibuat pada data uji. Hal ini memungkinkan untuk mengevaluasi kinerja model yang telah melalui proses pelatihan. Selanjutnya, node "Performance NB" digunakan untuk mengevaluasi kinerja model Naïve Bayes, menghasilkan matriks performa yang relevan. Proses ini sangat penting untuk memastikan bahwa model yang dibangun tidak hanya akurat tetapi juga mampu melakukan generalisasi dengan baik.



Gambar 9. Parameters Operator Split Data

Pada pengujian data ini, pada gambar 9 ratio pada parameter yang digunakan sebanyak 80:20 dengan dua kategori data yaitu, data latih sebanyak 80% (delapan puluh persen) dan data uji sebanyak 20% (dua puluh persen). Pada penggunaan algoritma Naive Bayes ini digunakan dataset sebanyak 1.393 data latih dan 322 data uji, agar menentukan seberapa banyak *accuracy* yang di dapatkan pada pengujian dataset pada peristiwa Konflik Palestina-Israel ini, dapat dilihat pada gambar 10, operator *Naive Bayes* digunakan untuk memproses sebuah data latih dan untuk operator *Apply Model* digunakan untuk mengalirkan dari data uji untuk mengevaluasi sebuah kinerja model yang sudah dilatih sebelumnya. *Operator performance* digunakan untuk mengukur nilai akurasi, prediksi dan *recall* pada data uji.

3.5 Confusion Matrix

Setelah melakukan proses implementasi algoritma, hasil dari dataset pengujian Naive Bayes Classifier mendapatkan sebuah confusion matrix yang dapat dilihat pada gambar 10, menghasilkan total 650 *True Positive* (TP), 402 total *True Negative* (TN), 0 total *False Positive* (FP), dan 61 total *False Negative* (FN). Dengan menggunakan algoritma Naive Bayes ini menghasilkan perhitungan sebagai berikut: *Accuracy* sebesar 94%, *Recall* sebesar 91%, *Precision* sebesar 100% dan *F1-Score* sebesar 95% dengan rumus perhitungan sebagai berikut:

accuracy: 94.52%

	true negatif	true positif
pred. negatif	402	61
pred. positif	0	650
class recall	100.00%	91.42%

Gambar 10. Confusion Matrix Algoritma Naive Bayes Classifier

$$Accuracy = A = \frac{TP + TN}{TP + TN + FP + FN} = \frac{(650 + 402)}{(650 + 402 + 0 + 61)} = \frac{1052}{1113} = 0.94 \quad (2)$$

$$Precision = \frac{TP}{TP + FN} = \frac{650}{650 + 61} = 0.91 \quad (3)$$

$$Recall = \frac{TP}{TP + FP} = \frac{650}{650 + 0} = 1 \quad (4)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.91 \times 1)}{(0.91 + 1)} = 2 \times \frac{0.91}{1.91} = 0.95 \quad (5)$$

Setelah mengimplementasikan algoritma Naive Bayes Classifier, hasil evaluasi pada gambar 10 menggunakan kumpulan data pengujian menghasilkan matriks konfusi yang merangkum performa model.

3.6 WordCloud

Setelah menghitung hasil nilai dari matriks *accuracy*, tahapan selanjutnya adalah menggunakan metode WordCloud yang dimana pada penelitian ini dibutuhkan untuk menghitung jumlah dari seberapa banyak kata yang sering muncul pada sebuah dataset, maka semakin besar ukurannya untuk muncul [29].



Gambar 11. Visualisasi Wordcloud Dataset Konflik Palestina-Israel



Gambar 11 menunjukkan visualisasi WordCloud dari kumpulan data mengenai konflik Palestina-Israel, di mana kata "Gaza" muncul dengan jumlah 972 kata dalam diskusi di media sosial X, menunjukkan bahwa isu terkait Gaza menjadi subjek utama percakapan dan perselisihan dalam kumpulan data tersebut. Selain itu, kata "Palestina" adalah kata kedua yang muncul sebanyak 850 kali dalam dataset, menunjukkan bahwa Palestina juga menjadi subjek yang signifikan dalam diskusi konflik. Penelitian menggunakan WordCloud untuk mengidentifikasi kata-kata kunci yang paling umum dan paling sering ditemukan dalam dataset. Hal ini membantu dalam pemahaman tren utama atau fokus pembicaraan publik terkait topik yang diteliti yaitu konflik Palestina-Israel.

4. KESIMPULAN

Berdasarkan hasil dari penelitian analisis sentimen dari penggunaan media sosial X tentang Konflik Palestina – Israel ini bertujuan untuk menganalisis sentimen warga negara Indonesia terhadap konflik Palestina-Israel yang ramai diperbincangkan di media sosial X (sebelumnya Twitter). Penelitian ini menggunakan algoritma pengklasifikasi Naive Bayes dan mengumpulkan data dari Oktober 2023 hingga Juni 2024. Dari 4.400 data asli yang dikumpulkan, 1.715 data relevan diperoleh setelah dilakukan penyaringan dan pra-pemrosesan. Pada tahap pelabelan, data diklasifikasikan menjadi kelas positif dan negatif dengan menggunakan total 1.393 data pelatihan dan 322 data pengujian. Proses pembobotan dilakukan dengan menggunakan metode TF-IDF (Term Frekuensi-Inverse Document Frekuensi) untuk menentukan bobot kata-kata penting. Hasil analisis menunjukkan bahwa model Naive Bayes Classifier memberikan kinerja yang sangat baik dengan akurasi 94%, presisi 91%, *recall* 100%, dan skor F1 95%. Model ini terbukti efektif dalam mengklasifikasikan sentimen pengguna media sosial X terkait konflik Palestina-Israel dengan akurasi tinggi dan analisis sentimen pada konflik Palestina-Israel ini memiliki lebih banyak tanggapan *positive* dikarenakan sebagian besar komentar menunjukkan dukungan yang kuat untuk Palestina. Hal tersebut dapat terlihat dari sentimen positif yang mengapresiasi terhadap bentuk ketahanan Palestina dalam menghadapi situasi konflik yang terus berjalan ini dan menghargai terhadap usaha masyarakat di Indonesia dalam mendukung Palestina. Oleh karenanya, penelitian ini mampu mengidentifikasi dan mengukur sentimen masyarakat Indonesia terhadap konflik Palestina-Israel dengan menggunakan algoritma pengklasifikasi Naive Bayes. Hal ini menunjukkan bahwa model ini dapat digunakan untuk analisis sentimen pada topik serupa di masa depan.

Silvia Damayanti Simamora - Analisis Sentimen Terkait Konflik Palestina – Israel Pada Media Sosial X Menggunakan Algoritma Naïve Bayes Classifier

ORIGINALITY REPORT

11%

SIMILARITY INDEX

10%

INTERNET SOURCES

5%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	eprints.uty.ac.id Internet Source	1%
2	docobook.com Internet Source	1%
3	Rani Yunita, Mia Kamayani. "Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi", Indonesian Journal of Computer Science, 2023 Publication	1%
4	dinastirev.org Internet Source	1%
5	www.stmik-budidarma.ac.id Internet Source	1%
6	ejournal.stmik-budidarma.ac.id Internet Source	1%
7	text-id.123dok.com Internet Source	<1%

8	Rihan Maulana, Apriade Voutama, Taufik Ridwan. "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC", Jurnal Teknologi Terpadu, 2023 Publication	<1 %
9	www.djournals.com Internet Source	<1 %
10	123dok.com Internet Source	<1 %
11	jurnal.ubl.ac.id Internet Source	<1 %
12	jurnalunibi.unibi.ac.id Internet Source	<1 %
13	jurnal.pcr.ac.id Internet Source	<1 %
14	Submitted to Universitas Brawijaya Student Paper	<1 %
15	siaiap32.univali.br Internet Source	<1 %
16	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %
17	ftsl.itb.ac.id Internet Source	<1 %

18	pdfcoffee.com Internet Source	<1 %
19	Eko Budi Santoso, Aryo Nugroho. "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook", Eksplora Informatika, 2019 Publication	<1 %
20	abdurrahmanpunya.blogspot.com Internet Source	<1 %
21	kc.umn.ac.id Internet Source	<1 %
22	www.slideshare.net Internet Source	<1 %
23	ejournal.itats.ac.id Internet Source	<1 %
24	tunasbangsa.ac.id Internet Source	<1 %
25	e-journal.my.id Internet Source	<1 %
26	Andreyestha Andreyestha, Agus Subekti. "ANALISA SENTIMENT PADA ULASAN FILM DENGAN OPTIMASI ENSEMBLE LEARNING", Jurnal Informatika, 2020 Publication	<1 %

27 Anindita Ekaning Saputri, Windia Hadi. <math><1\%</math>
"PENGEMBANGAN E-BOOK BERMUATAN
HIGH ORDER THINKING SKILL (HOTS)",
AKSIOMA: Jurnal Program Studi Pendidikan
Matematika, 2021
Publication

28 Dwi Rahma Firmansyah, Endang <math><1\%</math>
Lestariningsih. "Analisis Sentimen Ulasan
Aplikasi Smart Campus Unisbank di Google
Playstore Menggunakan Algoritma Naive
Bayes", Jurnal JTIC (Jurnal Teknologi Informasi
dan Komunikasi), 2024
Publication

29 archive.org <math><1\%</math>
Internet Source

30 djournals.com <math><1\%</math>
Internet Source

31 jisai.mercubuana-yogya.ac.id <math><1\%</math>
Internet Source

32 jurnal.unimed.ac.id <math><1\%</math>
Internet Source

33 pdfs.semanticscholar.org <math><1\%</math>
Internet Source

34 bufnets.tech <math><1\%</math>
Internet Source

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off