

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah metode campuran (*mixed methods*), dimana metode kuantitatif dan kualitatif digunakan secara bersamaan. Metode kuantitatif digunakan untuk mengevaluasi dan membandingkan kinerja dua algoritma, *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN), dalam analisis sentimen terkait konflik Palestina-Israel. Analisis ini meliputi akurasi pengukuran, presisi, recall, skor f1, skor MCC, *log loss*, dan waktu pemrosesan untuk kedua algoritma. Dengan menggunakan uji statistik, penelitian ini bertujuan untuk menentukan algoritma mana yang lebih efektif dalam mengklasifikasikan sentimen publik pada suatu topik dengan membuat perbandingan antara SVM dan CNN.

Pendekatan kualitatif dalam penelitian ini meliputi analisis konten untuk memahami perasaan dan pemikiran pengguna media sosial terkait konflik Palestina-Israel. Melalui analisis kualitatif, penelitian ini bertujuan untuk mendapatkan pemahaman lebih dalam mengenai opini publik terhadap konflik Palestina-Israel, memberikan wawasan tidak hanya berdasarkan statistik tetapi juga interpretasi teks yang lebih dalam. Dengan menggabungkan kedua metode tersebut, metode campuran ini tidak hanya mengukur dan membandingkan efektivitas algoritma, namun juga memberikan pandangan komprehensif tentang bagaimana opini publik yang diungkapkan dalam konteks konflik. Hal ini memungkinkan peneliti untuk lebih memahami dan memperkaya masalah penelitian.

3.2 Desain Penelitian

Desain yang digunakan dalam penelitian ini adalah *explanatory sequential mixed methods*, yang bertujuan untuk menggabungkan metode kuantitatif dan kualitatif untuk membandingkan kinerja dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN) dalam menganalisis konflik Palestina-Israel. Penelitian ini dimulai dengan tahap

kuantitatif untuk mengevaluasi efektivitas kedua algoritma, diikuti oleh tahap kualitatif untuk mendapatkan wawasan lebih mendalam tentang hasil analisis sentimen.

Berikut langkah-langkah umum yang akan peneliti gunakan dalam desain penelitian *explanatory sequential mixed methods* pada penelitian ini:

1. Pemilihan Sumber Data: Sumber data penelitian ini adalah tweet berbahasa Indonesia yang membahas konflik Palestina-Israel yang diperoleh dari jejaring sosial X.
2. Tahap Kuantitatif:
 - Persiapan Data: Melakukan prapemrosesan data termasuk pembersihan teks, tokenisasi, penghapusan stop word, dan normalisasi untuk memastikan data siap digunakan oleh algoritma. Proses ini penting untuk mengurangi noise dan meningkatkan kualitas data yang akan dianalisis.
 - Pengembangan Model: Mengimplementasikan model SVM dan CNN untuk melatih dan menguji dataset. Mengoptimalkan parameter SVM dan CNN untuk mendapatkan performa terbaik dalam analisis sentimen.
 - Evaluasi Performa: Menggunakan data pengujian untuk mengevaluasi performa kedua model pada metrik seperti akurasi, presisi, *recall*, *f1-score*, *MCC score*, *log loss*, dan waktu proses. Pengujian ini dilakukan untuk mengevaluasi seberapa baik setiap algoritma dalam mengidentifikasi emosi dari tweet.
3. Tahap Kualitatif:
 - Analisis Isi: Melakukan analisis isi untuk memahami konteks dan nuansa dari sentimen yang diungkapkan oleh pengguna media sosial terkait konflik Palestina-Israel.
4. Analisis Perbandingan:
 - Menganalisis dan membandingkan performa SVM dan CNN untuk menentukan algoritma mana yang lebih efektif dalam menganalisis sentimen terkait konflik Palestina-Israel. Hasil perbandingan ini

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

akan memberikan gambaran mengenai kelebihan dan kelemahan masing-masing algoritma.

5. Integrasi Hasil:

- Mengintegrasikan temuan dari kedua tahap untuk memberikan pemahaman yang komprehensif. Menggunakan hasil kuantitatif untuk menunjukkan efektivitas algoritma dan hasil kualitatif untuk memberikan konteks yang lebih dalam tentang sentimen publik.

6. Interpretasi dan Kesimpulan:

- Menarik kesimpulan dari hasil analisis perbandingan dan interpretasi kualitatif untuk membuat rekomendasi algoritma yang lebih baik dalam konteks analisis sentimen.

Desain *explanatory sequential mixed methods* dipilih dalam penelitian ini karena pendekatan ini memungkinkan peneliti untuk mengontrol variabel yang relevan, membandingkan efektivitas algoritma SVM dan CNN secara objektif, serta memberikan wawasan yang lebih kaya dan mendalam tentang sentimen publik terkait konflik Palestina-Israel. Dengan pendekatan ini, diharapkan dapat diperoleh hasil yang valid dan reliabel terhadap kinerja kedua algoritma dalam analisis sentimen serta pemahaman yang lebih mendalam tentang konteks sentimen yang dianalisis.

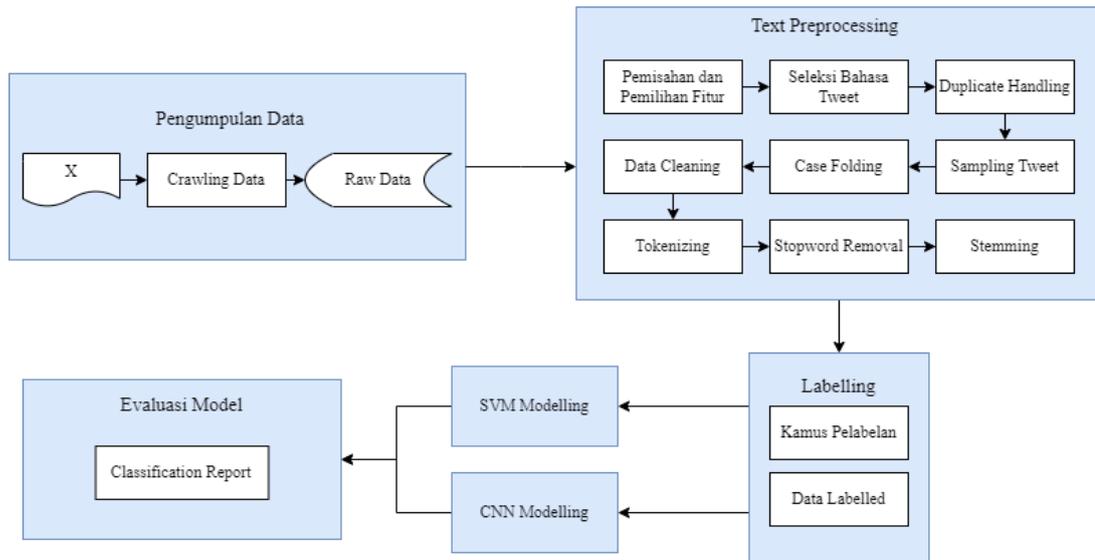
3.3 Prosedur Penelitian

Penelitian ini meliputi beberapa tahapan dan prosedur yang disesuaikan dengan desain penelitian yang digunakan yaitu *explanatory sequential mixed methods* proses ini dimulai dari tahap pengumpulan data dari X, kemudian dilanjutkan dengan *Text Preprocessing* yang mencakup berbagai teknik untuk mempersiapkan teks mentah menjadi siap untuk dianalisis. Selanjutnya, dilakukan pelabelan sentimen dengan menggunakan kamus pelabelan *lexicon*. Lalu penelitian ini berlanjut ke tahap *modelling*, yang mana menggunakan dua model *machine learning* yaitu *Support Vector Machine (SVM)* dan *Convolutional Neural Network (CNN)* dan terakhir melakukan evaluasi performa dari model yang telah dibangun seperti terlihat pada Gambar 3.1 berikut ini:

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu



Gambar 3.1 Tahapan Penelitian

3.3.1 Pengumpulan Data

Data yang akan diambil dan dikumpulkan adalah data *tweet* dari sosial media X yaitu berupa *text* yang menyatakan opini masyarakat terhadap konflik Palestina-Israel. Proses pengumpulan data ini disebut sebagai *crawling data* yaitu proses otomatis untuk mengumpulkan informasi dari berbagai sumber di internet termasuk di X. *Crawling* adalah pengumpulan data di situs web dengan memasukkan *Uniform Resource Locator (URL)*. URL ini adalah referensi untuk menemukan semua hyperlink di situs web. Kemudian pengindeksan dilakukan untuk menemukan kata-kata dalam dokumen pada setiap tautan yang ada. Untuk melakukan *crawling data* di X maka diperlukan akses ke X, semenjak *Twitter* berganti kepemilikan menjadi di bawah kepemimpinan Elon Musk dan X Corp, pengambilan data di X menggunakan API *Twitter* tidak lagi diberikan secara gratis, sehingga dibutuhkan cara lain untuk mengambil datanya. Dalam penelitian ini, pengumpulan data *tweet* memanfaatkan *tweet-harvest* yang dibuatkan dengan *Node.JS* dan dikembangkan oleh seorang YouTuber Indonesia bernama Helmi Satria. Secara garis besar pengambilan datanya tidak jauh berbeda dengan menggunakan API *Twitter*.

Dalam proses pengambilan datanya yang pertama dilakukan adalah memasukkan *keyword* dari *tweet* yang akan diambil dan rentang waktu *tweet* yang ingin diambil, dalam penelitian ini *keyword* yang digunakan adalah “Palestina”,

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

“Israel”, “Gaza”, dan “Hamam” rentang waktu *tweet* yang diambil adalah pada 1 Oktober 2023 sampai dengan 30 April 2024, waktu dipilih berdasarkan hangatnya isu ini pada rentang waktu tersebut, selanjutnya untuk menghubungkannya dengan X diperlukan *auth-token* yang dapat dilihat di setiap akun X, setelah itu hasil pengambilan data akan berbentuk file .csv yang berisi *tweet* dengan kata kunci yang dicari juga sesuai dengan rentang waktunya.

3.3.2 Text Preprocessing

Tahapan ini adalah tahap mengolah data mentah yang didapatkan dari hasil *crawling data* di X, menjadi data yang siap digunakan. Dalam *text preprocessing* terdapat beberapa tahapan yang juga digunakan dalam penelitian ini antara lain :

1. Pemisahan dan pemilihan fitur : Setelah data berhasil di *crawling*, data yang terambil belum dipisahkan menjadi kolom dan baris sesuai dengan fiturnya. Maka dari itu perlu dilakukan pemisahan setiap fitur yang didapat menjadi kolom tersendiri, selain itu dalam penelitian ini hanya akan menggunakan 2 fitur saja yaitu *created_at* atau waktu *tweet* di posting dan *full_text* atau isi dari *tweet*, maka dari itu dalam proses ini pula akan dilakukan penghapusan fitur yang tidak perlu dan mengubah format *created_at* menjadi *datetime* agar mudah diproses nantinya.
2. Seleksi bahasa *tweet* : Pada tahap ini data yang telah berhasil dipisahkan fiturnya akan di seleksi berdasarkan bahasa *tweet* yang digunakan. Sesuai dengan Batasan masalah dalam penelitian ini, *tweet* yang dianalisis hanya *tweet* berbahasa Indonesia, maka dari itu *tweet* yang berbahasa selain bahasa Indonesia akan dihapus.
3. *Duplicate Handling* : Tahapan ini merupakan tahapan untuk mengatasi adanya data duplikat dari data *tweet* yang telah diambil. Tentunya data duplikat akan membuat redundansi data dan ketidakefektifan hasil analisis karena mengolah data yang sama lebih dari 1. Dalam proses ini *tweet* yang berisi kalimat yang sama akan dihilangkan dan disisakan 1 saja.
4. *Sampling tweet* : Pada tahap ini, akan diambil sampel *tweet* dengan jumlah yang sama setiap bulannya selama tujuh bulan, hal ini dilakukan agar proporsi *tweet* sama setiap bulannya sehingga dapat dilihat angka-angka

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

yang menarik dalam hasil sentimen nantinya, misalnya apakah ada kenaikan sentimen positif atau negatif dan lain sebagainya.

5. *Case Folding* : Tahap ini bertujuan mengubah semua huruf dalam teks dari data yang diambil, menjadi huruf kecil atau huruf besar. Selain mengubah huruf, karakter lain yang bukan huruf dan angka, seperti tanda baca atau spasi dianggap sebagai *delimiter*. *Delimiter* ini dapat diabaikan atau dihapus. Tujuan dari *Case Folding* adalah untuk mencegah kekeliruan dalam pemrosesan kata dan menyeragamkannya sehingga hasilnya sama, karena bisa terdapat perbedaan hasil jika kalimat yang sama ditulis dengan huruf kapital dengan huruf kecil.
6. *Data cleaning* : Pada bagian cleaning kalimat yang ada dibersihkan dari adanya komponen yang dapat mengganggu proses analisis yang disebut *delimiter* tadi, contohnya tanda baca, URL, *username*, *Retweet*, Karakter HTML, dan *hashtag*.
7. *Tokenizing* : Proses ini akan memisahkan teks yang didapat menjadi kata per kata, atau unit-unit kecil yang disebut dengan token.
8. *Stopword Removal* : Tahapan akan melakukan penghapusan pada kata-kata yang muncul dalam stoplist, kata yang dihapus biasanya adalah kata penghubung yang tidak relevan dengan apa yang Tengah diteliti contohnya seperti kata penghubung dan, maka, tetapi dan lain sebagainya
9. *Stemming* : *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk kata menjadi bentuk dasarnya . Proses pemetaan dan penguraian digunakan untuk menemukan kata dasar dari sebuah kata yang mengalami imbuhan dengan cara menghilangkan atau menghapus imbuhan-imbuhan tersebut.

3.3.3 Labelling

Labelling merupakan pemberian nilai sentimen terhadap suatu teks yang dapat berupa positif, negatif dan netral. Pada penelitian ini *labelling* dilakukan dengan memanfaatkan kamus pelabelan *Lexicon Based Features* yang mana kamus ini berisi kumpulan kata dengan berbagai nilai yang merepresentasikan sentimen dari kata-kata yang ada, baik positif, negatif maupun netral.

Maulana Wirayudha, 2024

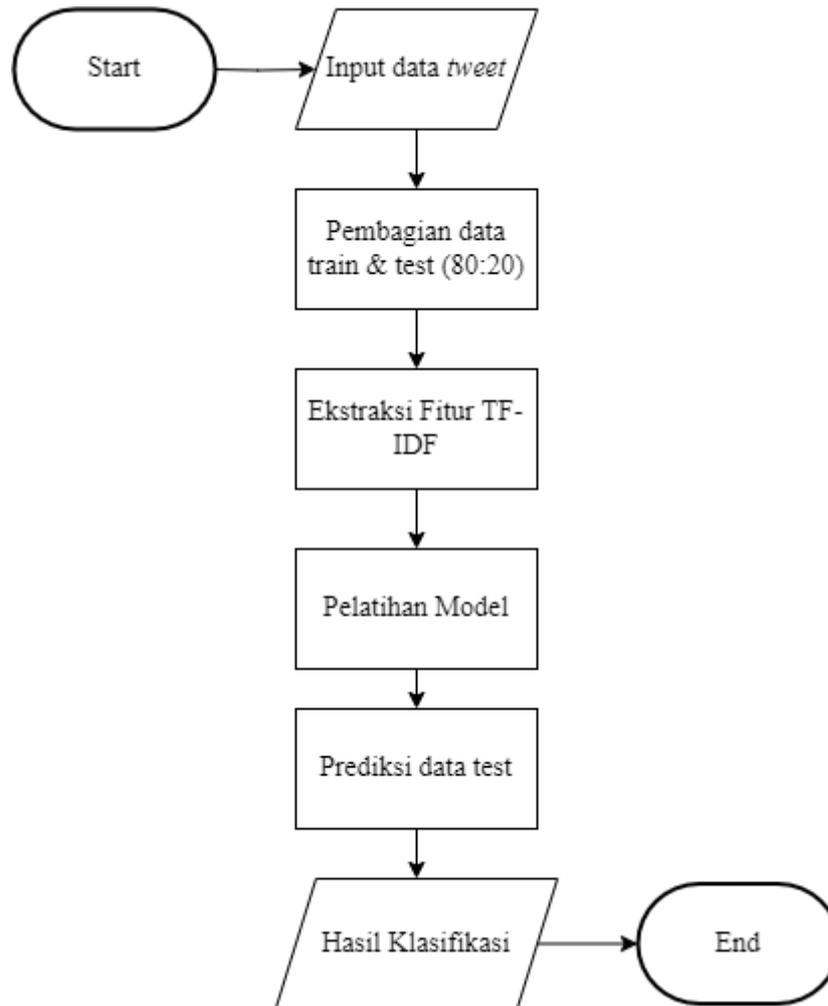
PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

3.3.4 Modelling

Setelah tahapan labelling dilakukan, selanjutnya masuk ke tahap modelling. Pada tahap ini, data yang telah dilabeli akan diuji menggunakan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN). Pemodelan dilakukan untuk melihat seberapa akurat model dalam mengklasifikasikan data *tweet* berdasarkan sentimennya ke dalam tiga jenis sentimen, yaitu positif, negatif, dan netral. Pemodelan menggunakan dua algoritma dilakukan untuk melihat perbedaan hasil antara kedua model yang digunakan.

Modelling dengan menggunakan algoritma *Support Vector Machine* (SVM) dibuat dengan konfigurasi pembagian data menjadi 80:20 yaitu 80% data latih dan 20% data uji, vektorisasi teks yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) dan kernel linear.



Gambar 3.2 Diagram alir algoritma SVM

Pada Gambar 3.2, terdapat ilustrasi dari algoritma SVM model 1 yang dimulai dari memasukkan data *tweet* yang telah dilabeli. Pertama, data *tweet* dibagi menjadi 80:20, yaitu 80% sebagai data train dan 20% sebagai data test. Selanjutnya, proses masuk ke tahap ekstraksi fitur dengan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah data *tweet* menjadi vektor. Setelah itu, dilakukan pelatihan model dengan menggunakan data train. Pada tahap selanjutnya, model akan diterapkan untuk memprediksi data test, dan hasil dari prediksi tersebut akan menampilkan *output* berupa *Confusion Matrix* dan *Classification Report* untuk dianalisis lebih lanjut.

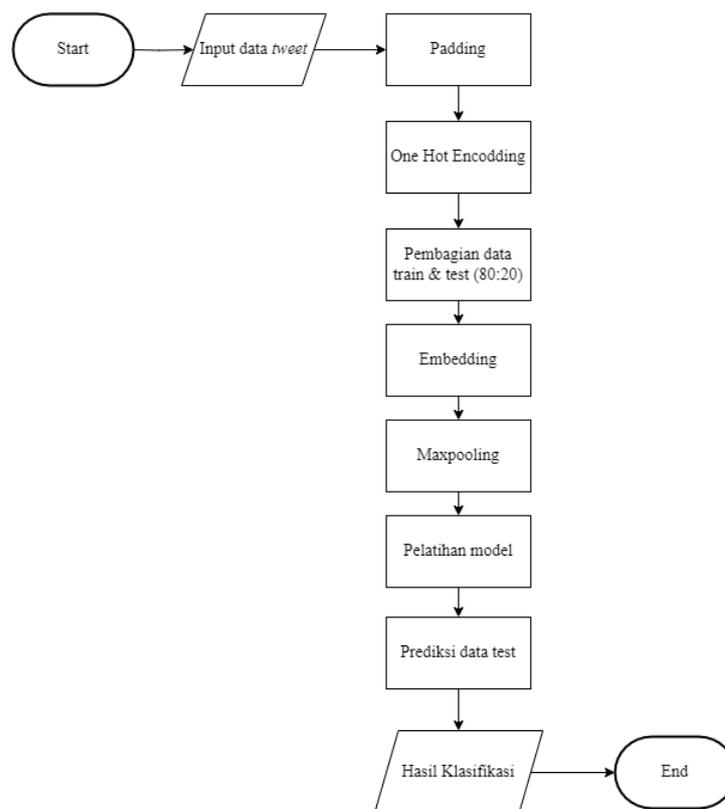
Untuk pemodelan dengan menggunakan algoritma CNN, dibuat model dengan konfigurasi pembagian data 80:20 yaitu 80% data latih dan 20% data uji,

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

max_words = 1000, activation = relu, Dense = 3, optimizer = adam, epochs = 10 dengan menggunakan model CNN 1 Dimensi atau 1D CNN.



Gambar 3.3 Diagram alir algoritma CNN

Pada Gambar 3.3 mengilustrasikan alur algoritma CNN model 1. Pada model ini dimulai dari memasukkan data *tweet*, lalu menambahkan *padding*. *Padding* ini adalah masukkan kosong yang diisi untuk menyamakan ukuran panjang setiap *tweet*. Contohnya, pada model ini didefinisikan panjang maksimal kata dari setiap *tweet* adalah 1000, maka *tweet* yang tidak berisi 1000 kata akan ditambahkan hingga berjumlah 1000 dengan masukkan kosong ini yang disebut pad. Setelah dilakukan *padding*, dilanjutkan dengan melakukan *one hot encoding*, yaitu teknik mengubah data kategorikal, yaitu kata dalam *tweet*, agar menjadi format yang dapat dimasukkan ke dalam model ini.

Setelah membagi data menjadi data pelatihan (train) dan data uji (test) dengan perbandingan 80:20 seperti pada model-model sebelumnya, tahap berikutnya dalam model CNN adalah memasukkan data ke dalam *embedding layer*. *Embedding layer* bertugas untuk mengubah teks dari *tweet* menjadi representasi

Maulana Wirayudha, 2024

PERBANDINGAN PERFORMA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA ANALISIS SENTIMEN KONFLIK PALESTINA ISRAEL DI X

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

vektor, mirip dengan proses TF-IDF yang digunakan dalam model SVM. Proses *embedding* ini sangat penting karena memungkinkan model untuk memahami hubungan dan makna antara kata-kata dalam *tweet*. Setelah melakukan *embedding*, nilai vektor untuk setiap kata akan ditemukan, dan kemudian dicari kata dengan nilai vektor tertinggi dalam setiap *tweet*. Hal ini disebut dengan proses *maxpooling* (*maxpool*). Dengan menggunakan nilai vektor tertinggi ini, model dapat menangkap informasi kunci atau fitur penting dari setiap *tweet* yang akan digunakan dalam pelatihan model

Setelah melalui tahapan *embedding* dan *maxpool*, model CNN akan dilatih menggunakan data train untuk mengenali pola dan fitur yang ada dalam teks *tweet* terkait konflik Palestina-Israel. Proses pelatihan ini bertujuan untuk mengoptimalkan bobot dan parameter dalam model sehingga dapat menghasilkan prediksi sentimen yang akurat pada data test. Setelah pelatihan selesai, dilakukan evaluasi model menggunakan data test untuk menghasilkan *classification report* dan *confusion matrix* yang memberikan gambaran tentang performa model dalam mengklasifikasikan sentimen *tweet* menjadi positif, negatif, atau netral. Hasil evaluasi ini nantinya dapat memberikan insight yang berguna dalam pemahaman sentimen publik terkait konflik tersebut.

3.3.5 Evaluasi

Dalam penelitian ini evaluasi performa dilakukan dengan menggunakan fungsi klasifikasi, dengan menggunakan fungsi *Classification Report* di mana yang dilihat nantinya adalah hasil evaluasi berupa nilai *Accuracy*, *Precision*, *Recall* dan *F1-Score* keempat metrik evaluasi tersebut didapat dengan menghitung jumlah prediksi benar dan salah dalam prediksi positif maupun negatif yaitu, *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang dirangkum dalam *confusion matrix*. Selain itu alat ukur lain dalam evaluasi yang akan digunakan juga adalah dengan menghitung nilai *Matthew Correlation Coefficient* (MCC) dan *Logarithmic Loss*

3.4 Lingkungan Komputasi

Penelitian ini didukung dengan beberapa alat pendukung baik itu perangkat keras maupun perangkat lunak, berikut merupakan lingkungan komputasi yang digunakan dalam penelitian ini.

Perangkat Keras:

1. Asus X441BA
2. RAM 8GB
3. SSD 512 GB
4. AMD A9
5. VGA AMD Radeon R5
6. Mouse Eksternal

Perangkat Lunak :

7. *Google Collaboratory*
8. X
9. Bahasa Pemrograman *Python*
10. Windows 10