

## BAB 3 METODOLOGI PENELITIAN

### 3.1 Metodologi Penelitian

Proses pembuatan Analisis Sentimen Twitter untuk pertandingan Indonesia vs Qatar di Piala Asia U-23 2024 menggunakan Support Vector Machine (SVM) melibatkan beberapa langkah penting:

#### 1. Studi Literatur

Penelitian ini dimulai dengan melakukan studi literatur. Tujuannya adalah untuk memperoleh pemahaman yang mendalam mengenai subjek yang akan diteliti. Informasi yang diperoleh diambil dari sumber-sumber seperti jurnal ilmiah, tulisan-tulisan terkait, dan buku-buku. Penelitian ini mengacu pada beberapa teori, antara lain mengenai pertandingan antara Indonesia dan Qatar dalam Piala Asia U-23 2024, analisis sentimen, metode TF-IDF, Support Vector Machine, dan Confusion Matrix.

#### 2. Pengumpulan Data

Hal kedua yang dilakukan adalah pengambilan data dari Twitter menggunakan tools tweets harvests. tools yang digunakan untuk melakukan crawling data pada media sosial Twitter dengan Application Programming Interface (API) menggunakan Node.js.

#### 3. Perancangan

Tahap awal perancangan sistem diawali dengan pembuatan diagram alur (flowchart) yang menggambarkan proses penelitian, pelabelan, praproses data, serta pembagian data menjadi set latihan dan set uji.

#### 4. Implementasi Algoritma

Setelah menyelesaikan tahap perancangan, langkah berikutnya adalah menerapkan pembobotan TF-IDF, lalu melakukan klasifikasi dengan SVM.

#### 5. Pengujian

Pada tahap ini, dilakukan pengujian perbandingan performa menggunakan 3 kernel pada algoritma SVM yaitu *kernel linear*, *kernel polynomial*, dan *kernel rbf* dengan menggunakan rasio pembagian data pelatihan dan pengujian yang berbeda, yaitu 60:40, 70:30, dan 80:20. Hal ini dilakukan untuk mengetahui

kombinasi kernel dan rasio data yang menghasilkan performa terbaik. Metrik yang digunakan dalam mengukur performa sistem adalah tingkat akurasi, presisi, recall, dan f1-score. Perhitungan metrik ini dilakukan menggunakan confusion matrix, yang menunjukkan hasil klasifikasi sistem secara detail.

## 6. Penulisan Laporan

Laporan ini disusun dengan tujuan untuk mendokumentasikan seluruh proses penelitian, perancangan, dan pengembangan aplikasi sampai selesai, sehingga dapat menjadi sumber informasi yang bermanfaat untuk para peneliti dengan studi yang sama.

## 7. Konsultasi

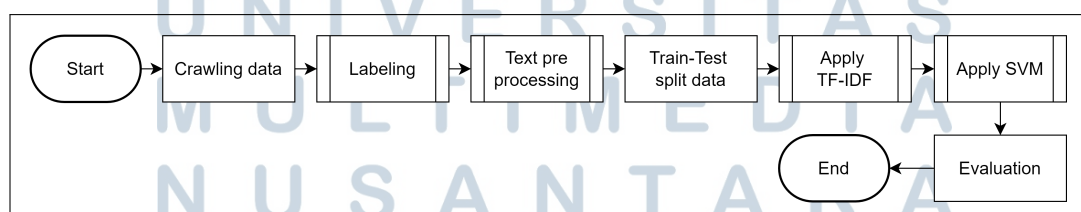
Tujuan utama konsultasi dengan dosen pembimbing adalah untuk mencari masukan dan saran untuk meningkatkan kualitas aplikasi dan laporan penelitian.

### 3.2 Perancangan Sistem

Pada tahap ini akan memaparkan skema *flowchart* sistem yang telah dibangun. Skema tersebut meliputi *flowchart* perancangan umum, *flowchart* pra-pemrosesan, *flowchart* penerapan TF-IDF, dan alur kerja penerapan mesin vektor dukungan.

#### 3.2.1 Gambaran Umum Perancangan

Perancangan secara umum melibatkan serangkaian langkah yang tergambar pada penelitian, yang direpresentasikan dalam bentuk diagram alur atau Flowchart, yang menggambarkan proses dari awal perancangan selesai.



Gambar 3.1. Flowchart Gambaran Umum Perancangan

Alur perancangan sistem yang digambarkan pada Gambar 3.1 dimulai

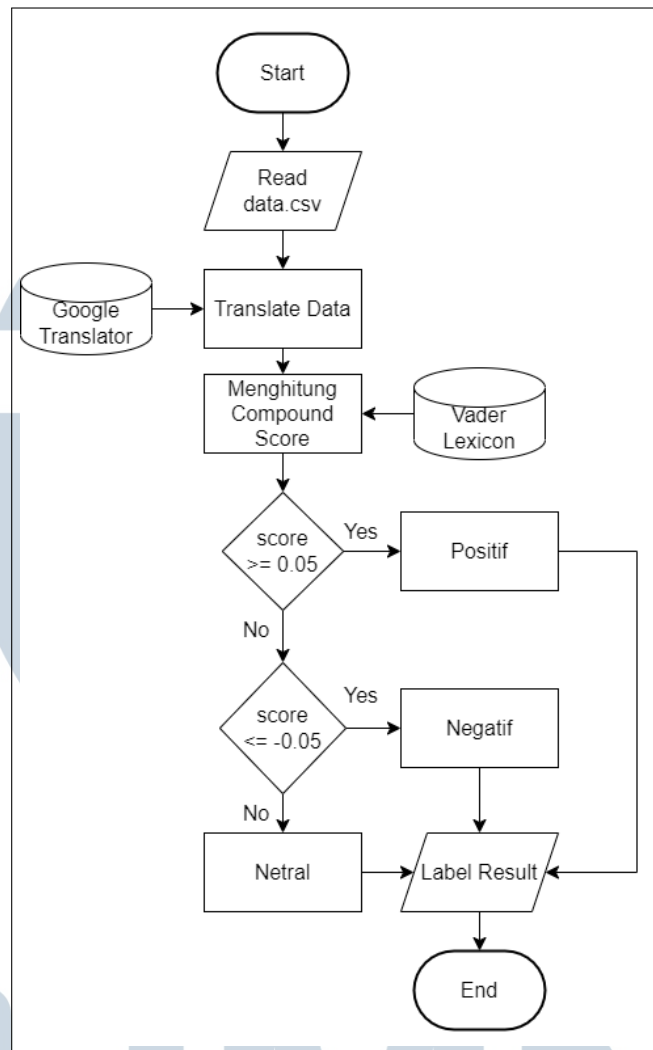
dengan *crawling data, labeling, Text pre-processing, Train-Test split data, Apply TF-IDF, Apply SVM, Evaluation*

### **3.2.2 Crawling Data**

Mengambil informasi dari sebuah situs web menggunakan program otomatisasi dan API sebagai saluran komunikasi adalah teknik yang dikenal sebagai "crawling data" [20]. Dalam penelitian ini, teknik tersebut diterapkan menggunakan tools tweets harvests untuk mengekstrak tweet dari sejumlah pengguna dengan topik penelitian yang ditentukan. Fokus pencarian adalah pada kata kunci "pertandingan Indonesia vs Qatar di Piala Asia U-23" dan hanya tweet dalam bahasa Indonesia yang diambil dari tanggal 15 April 2024 hingga 26 April 2024. Rentang waktu tersebut dipilih untuk mencakup periode setelah pertandingan selesai hingga Timnas Indonesia mencapai babak semifinal, menghasilkan total 2258 tweet yang berhasil diambil.

### **3.2.3 Labeling Data**

Data labeling adalah proses yang memasukkan sentimen ke dalam tweet yang telah diambil melalui pengambilan data. Pada langkah pemberian label ini, digunakanlah alat yang dikenal dengan nama Valence Aware Dictionary and Sentiment Reasoner (VADER). VADER dirancang untuk menganalisis sentimen dalam teks dengan menggunakan aturan yang disesuaikan untuk memahami ekspresi emosi. Dalam penggunaannya, VADER menggabungkan kamus sentimen yang terdiri dari daftar kata-kata leksikal yang umumnya diberi label sebagai kata-kata positif, negatif, dan netral. Dalam proses pelabelan menggunakan VADER, jika skor compound  $\geq 0.05$  maka sentimen dikategorikan sebagai positif; jika skor  $\leq -0.05$  maka sentimen dikategorikan sebagai negatif; dan jika skor berada di antara  $> -0.05$  dan  $< 0.05$  maka sentimen dikategorikan sebagai netral [21].

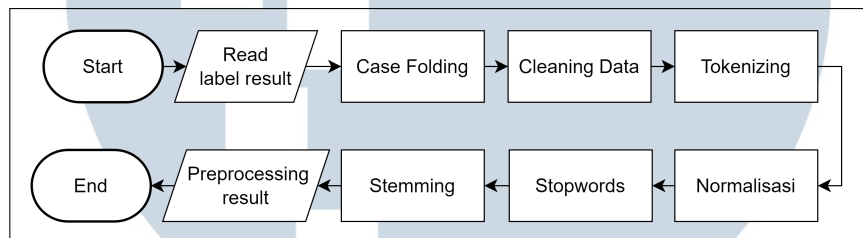


Gambar 3.2. Flowchart Labeling Data

Gambar 3.2 menunjukkan langkah-langkah pada proses pelabelan data. Pertama, data hasil crawling dibaca. Selanjutnya, data diterjemahkan ke bahasa Inggris menggunakan *Library Google Translator* karena VADER hanya mendukung bahasa Inggris. Setelah diterjemahkan, skor sentimen dihitung menggunakan VADER. Jika skor  $\geq 0.05$ , sentimennya positif. Jika skor  $\leq -0.05$ , sentimennya negatif. Jika skor  $> -0.05$  dan  $< 0.05$ , sentimennya netral.

### 3.2.4 Text Preprocessing

*Text Preprocessing* adalah sebuah langkah penting dalam mengubah data teks menjadi lebih teratur dan terstruktur. Sebuah penelitian oleh [22] menemukan bahwa urutan terbaik dalam melakukan persiapan data teks adalah dengan melakukan normalisasi terlebih dahulu, sebelum dilanjutkan ke tahap pembuangan kata berhenti (stopword removal) dan stemming. Seperti yang ditunjukkan pada Gambar 3.3, persiapan data teks terdiri dari beberapa langkah, yaitu *text preprocessing* terdiri dari beberapa tahapan yaitu *casefolding*, *cleaning data*, *tokenizing*, *normalisasi*, *stopwords removal*, dan *stemming*.



Gambar 3.3. Flowchart Text Preprocessing

#### 1. Case Folding

Pada tahap ini, semua huruf dalam teks, baik huruf besar maupun huruf kecil, akan diubah menjadi huruf kecil. Hal ini dilakukan untuk memudahkan proses pengolahan teks selanjutnya.

#### 2. Cleaning Data

Pembersihan teks dilakukan pada tahap ini, dengan menyingkirkan semua elemen yang tidak termasuk dalam data teks yang ingin dianalisis, seperti *username*, *hashtag*, *simbol*, *emotikon*, *tautan*, *karakter tunggal*, *angka*, *spasi*, dan *tanda baca*.

#### 3. Tokenizing

*tokenizing* adalah proses memecah kalimat menjadi kata-kata individual. Langkah ini merupakan langkah awal untuk mempersiapkan tahap selanjutnya, yaitu normalisasi, penghapusan kata berhenti, dan pemotongan kata dasar.

#### 4. Normalisasi

Normalisasi dilakukan untuk membakukan kata-kata yang tidak resmi agar

sesuai dengan standar bahasa yang berlaku. Hal ini dilakukan dengan mengubah kata-kata singkatan, slang, dan dialek menjadi bentuk yang baku.

#### 5. *Stopwords*

Proses ini bertujuan untuk mengidentifikasi kata-kata yang paling penting pada teks serta menghilangkan kata-kata yang tidak memiliki arti yang signifikan, seperti kata sambung.

#### 6. *Stemming*

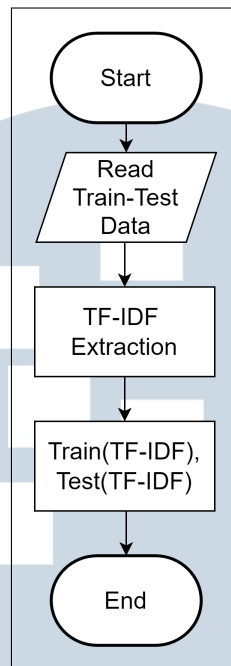
Dalam proses stemming, awalan dan akhiran kata dihilangkan untuk mendapatkan kata dasarnya.

### 3.2.5 *Train-Test Split Data*

Pada bidang machine learning, teknik *Train-Test Split Data* digunakan untuk memisahkan data menjadi dua bagian yaitu *data training* dan *data testing*. Tujuannya adalah untuk mengevaluasi performa model machine learning. Pembagian ini dilakukan sebelum proses pembobotan, sehingga data yang dipecah masih dalam bentuk teks. Hal ini memungkinkan kita untuk membaca dan membandingkan label kalimat sebelum dan sesudah penerapan SVM.

### 3.2.6 *Penerapan TF-IDF*

Pembobotan TF-IDF merupakan langkah penting yang harus dilakukan sebelum menjalankan model klasifikasi dengan metode *support vector machine* untuk menghasilkan performa terbaik. Pembobotan ini penting untuk memberikan nilai yang tepat pada setiap kata dalam data, sehingga algoritma *support vector machine* dapat menginterpretasikan input dengan baik. Setelah membagi data menjadi Train dan Test, langkah selanjutnya adalah melakukan pembobotan kata menggunakan metode TF-IDF. Hasilnya adalah data *Train (TF-IDF)* dan *Test (TF-IDF)*, seperti yang ditunjukkan pada Gambar 3.4.

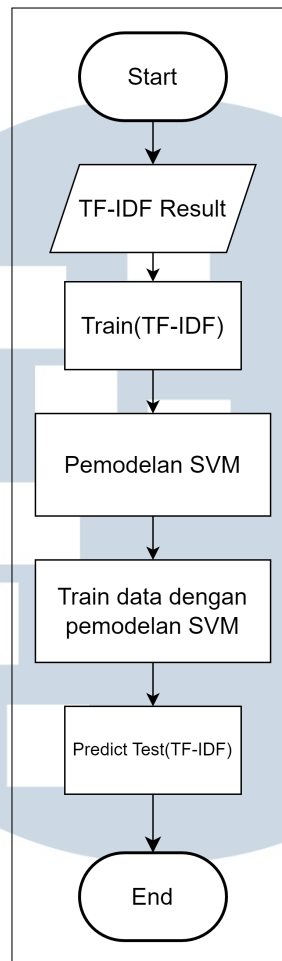


Gambar 3.4. Flowchart Apply TF-IDF

### 3.2.7 Penerapan SVM

Setelah pembobotan teks menggunakan TF-IDF, metode Support Vector Machine digunakan untuk mengklasifikasikannya. Seperti yang ditunjukkan pada Gambar 3.5, model yang dilatih pada data Train dengan TF-IDF digunakan untuk memprediksi label data Test yang juga telah dibobot dengan TF-IDF.

UIN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Gambar 3.5. Flowchart Apply SVM

### 3.2.8 Evaluation

Untuk mengevaluasi performa algoritma Support Vector Machine dalam mengklasifikasikan data, Confusion Matrix digunakan pada tahap evaluasi. Matriks ini mencakup kategori positif, netral, dan negatif. Nilai True menunjukkan kecocokan antara prediksi sentimen sebuah teks dengan label sebenarnya, sedangkan nilai False menunjukkan ketidakcocokan. Hasil Confusion Matrix kemudian digunakan untuk menghitung metrik evaluasi seperti Akurasi, Presisi, Recall, dan F1-score.