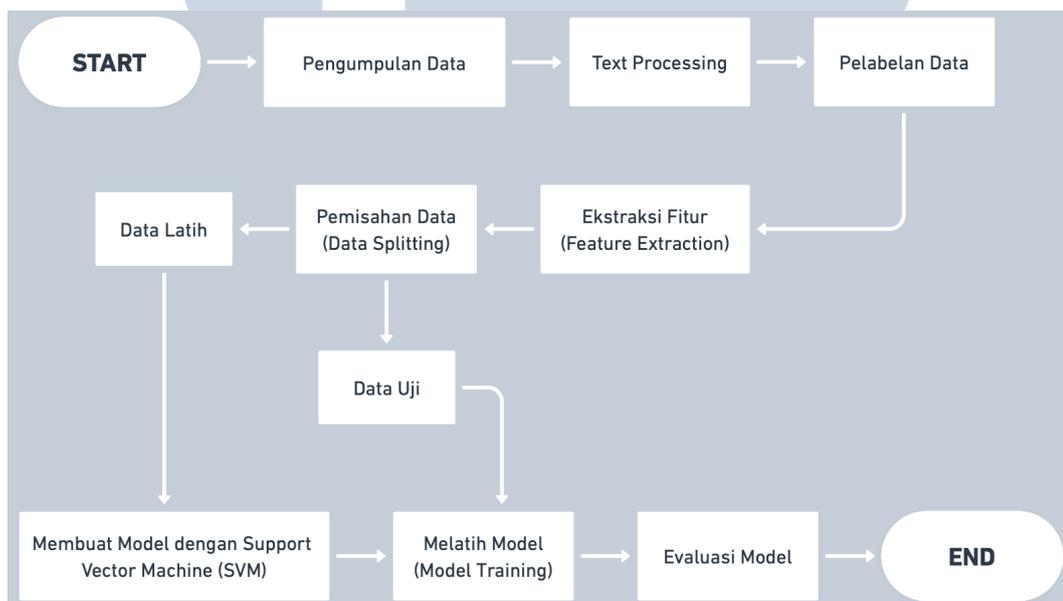


BAB 3 METODOLOGI PENELITIAN

3.1 Alur Diagram Penelitian

Pada tahapan ini dijabarkan langkah-langkah yang hendak dilakukan dalam menyusun dan mengerjakan penelitian. Beberapa tahapan yang akan dilakukan antara lain Pengumpulan Data (*Twitter scraping tools* dengan *Tweet-Harvest*), *Text Preprocessing*, Pelabelan Data (*Labeling Data*), Ekstraksi Fitur (*Feature Extraction*), Pemisahan Data (*Data Splitting*), Membuat Model dengan *Support Vector Machine (SVM)*, Melatih Model *Model Training*, dan Evaluasi Model. Tahapan tersebut digambarkan melalui *flowchart* pada gambar 3.1.



Gambar 3.1. Struktur metodologi penelitian

Berdasarkan Gambar 3.1, penjelasan mengenai setiap langkah yang akan dilakukan dapat diikuti sebagai berikut:

1. Pengumpulan Data

Proses ini akan dilakukan menggunakan *scraping tool* bernama *Tweet-Harvest* dengan dua kata kunci, yaitu "Polri" dan "Kinerja Polri" dengan tanggal pengambilan data dari tanggal 1 Februari 2022 hingga tanggal 1 Februari 2023. Pemilihan kata kunci dalam penelitian tersebut dilakukan berdasarkan konteks yang relevan dengan topik dan untuk lebih fokus pada

Polri dan kinerjanya. Kedua kata kunci tersebut dipilih berdasarkan seberapa sering kata-kata tersebut muncul dan seberapa kuat pengaruh emosionalnya dalam teks. Kata-kata yang sering muncul dan memiliki tingkat emosi yang kuat biasanya memberikan petunjuk yang lebih jelas mengenai sentimen.

2. *Text Processing*

Text Processing dilakukan untuk menganalisis dan menyortir data teks yang tidak terstruktur agar mendapatkan informasi yang diinginkan. Berikut merupakan penjelasan rinci dari tahap *Text Processing*:

- *Case Folding*, mengubah huruf kapital (*upper case*) menjadi huruf kecil (*lower case*).
- *Data Cleaning*, menghapus *hashtag*, *emoji*, *mention*, *username*, dan URL dari isi tweet.
- *Tokenization*, memisahkan kalimat menjadi daftar kata tunggal.
- *Normalization*, mengembalikan kata-kata tidak baku menjadi kata dasarnya.
- *Stopword*, menghapus kata-kata yang tidak penting menggunakan algoritma *stoplist*.
- *Stemming*, mereduksi kata-kata menjadi bentuk dasar dengan menghilangkan imbuhan.

3. Pelabelan Data (*Labeling Data*)

Pelabelan Data (*Data labeling*) dilakukan untuk memberikan *value* positif dan negatif dari tweet pengguna Twitter yang telah dikumpulkan pada langkah pengumpulan data.

4. Ekstraksi Fitur (*Feature Extraction*)

Ekstraksi Fitur (*Feature Extraction*) dilakukan untuk mengubah sekumpulan data yang besar menjadi fitur-fitur yang relevan.

5. Pemisahan Data (*Data Splitting*)

Pemisahan Data (*Data Splitting*) dilakukan dengan cara membagi data menjadi beberapa bagian agar model dapat dilatih dan diuji.

6. Membuat Model dengan *Support Vector Machine* (SVM)

Proses pembuatan model dilakukan pada data latih yang nantinya model ini akan digunakan data uji.

7. Melatih Model (*Model Training*)

Melatih Model *Model Training* dilakukan untuk menemukan bobot dan bias terbaik dalam suatu algoritma untuk meminimalkan fungsi error pada rentang prediksi. Fungsi error ini menentukan cara untuk mengoptimalkan algoritma machine learning.

8. Evaluasi Model

Evaluasi Model akan dilakukan menggunakan *confusion matrix* untuk menilai performa pada klasifikasi dalam *machine learning*. Metrik performa yang akan dinilai adalah Akurasi (*Accuracy*), *Recall*, Presisi (*Precision*), dan *F1 Score*.

3.2 Pengumpulan Data

Pengumpulan data adalah proses otomatis yang mengumpulkan dan mengindeks informasi dari berbagai sumber, seperti situs web, *database*, atau dokumen. Salah satu cara untuk mengumpulkan data adalah melalui media sosial Twitter dengan menggunakan *tool* bernama Tweet Harvest. Keunggulan Tweet Harvest adalah dapat dijalankan melalui *Command Line Interface* (CLI), seperti Command Prompt atau Terminal, dan hanya memerlukan *user valid auth_token*. Penggunaan Tweet Harvest tergambar secara visual dalam diagram alur yang disajikan pada Gambar 3.2.



Gambar 3.2. Proses pengumpulan data

Berdasarkan Gambar 3.2, dijelaskan langkah-langkah yang dilakukan dalam proses pengumpulan data menggunakan Tweet Harvest sebagai berikut:

1. Proses Instalasi Tweet-Harvest

Proses ini dilakukan pada CLI, yaitu Terminal.

2. *Input* Tanggal Tweet Diambil

Pada proses ini, tanggal tweet yang akan diambil mulai dari tanggal 1 Februari 2022 hingga 1 Februari 2023. Pembagian pencarian dilakukan dengan membagi setiap pencarian menjadi interval waktu selama 4 bulan.

3. *Input* Kata Kunci

Dalam proses ini, terdapat dua kata kunci yang digunakan, yaitu "Kinerja Polri" dan "Polri". Pencarian kata kunci ini dilakukan secara terpisah karena dalam penggunaan Tweet Harvest, hanya satu kata kunci yang dapat dimasukkan dalam satu pencarian.

4. *Input* Twitter Auth Token

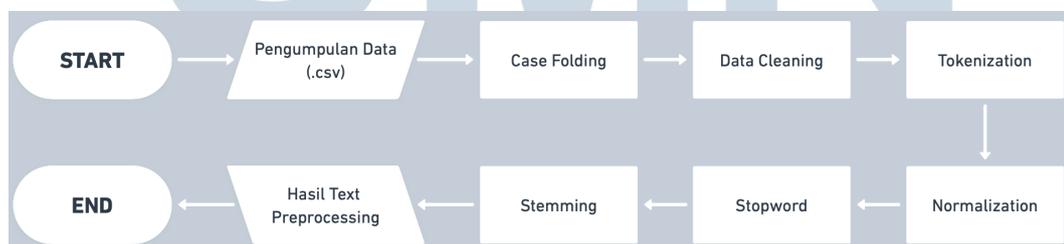
Untuk menggunakan fitur pencarian pada media sosial Twitter, Tweet Harvest memerlukan *User Auth Token* yang sudah melakukan *Login* terlebih dahulu.

5. Proses *Scraping Data* Twitter

Proses ini akan dilakukan secara otomatis berdasarkan input yang sudah ditentukan diatas. Setelah proses ini selesai maka *output* yang diperoleh dari data yang diambil akan langsung disajikan dalam bentuk *file Comma-Separated Values (CSV)* [15].

3.3 Text Preprocessing

Tahap ini bertujuan untuk menghilangkan *noise* dalam *tweet* agar memudahkan perhitungan bobot dan analisis pada tahap berikutnya, sehingga diharapkan hasil klasifikasi menjadi lebih akurat [2]. Prosesnya tergambar pada diagram alur Gambar 3.3.



Gambar 3.3. Proses *text preprocessing*

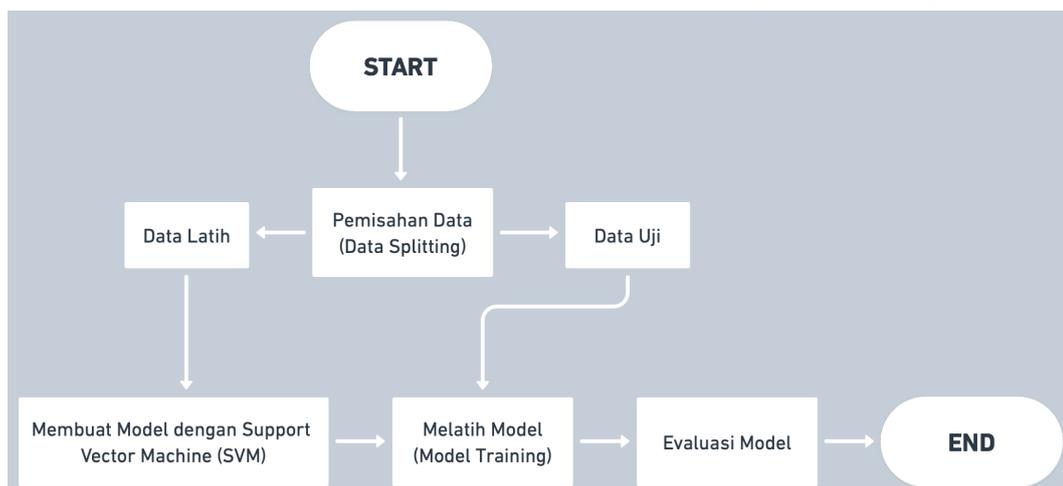
Langkah-langkah yang dilakukan pada tahap ini meliputi:

1. *Case Folding*: mengubah huruf kapital (*upper case*) menjadi huruf kecil (*lower case*).

2. *Data Cleaning*: menghapus *hashtag*, *emoji*, *mention*, *username*, dan URL dari isi *tweet*.
3. *Tokenization*: memisahkan kalimat menjadi daftar kata tunggal.
4. *Normalization*: mengembalikan kata-kata tidak baku menjadi kata dasarnya.
5. *Stopword*: menghapus kata-kata yang tidak penting menggunakan algoritma *stoplist*.
6. *Stemming*: mereduksi kata-kata menjadi bentuk dasar dengan menghilangkan imbuhan.

3.4 Analisis dengan Support Vector Machine (SVM)

Algoritma SVM digunakan untuk menganalisis data karena memiliki kemampuan efektif dalam memisahkan batas yang jelas, terutama dalam ruang dimensi yang tinggi [1]. Sebelum menerapkan metode tersebut, data diuji dan dilatih. Selanjutnya, dilakukan pengujian data menggunakan metode SVM seperti yang ditunjukkan pada Gambar 3.4. Gambar tersebut menampilkan alur kerja metode SVM dengan menggunakan *dataset* yang sudah melalui proses Pelabelan Data.



Gambar 3.4. Proses analisis data menggunakan *Support Vector Machine* (SVM)