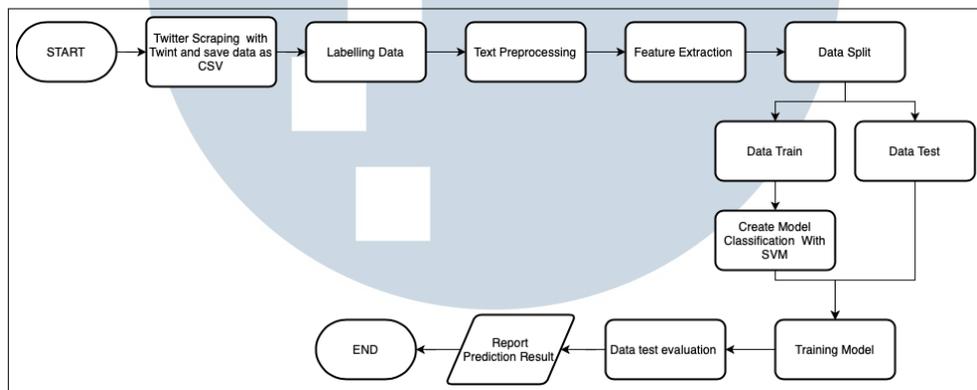


BAB 3 METODOLOGI PENELITIAN

3.1 Alur Diagram Penelitian

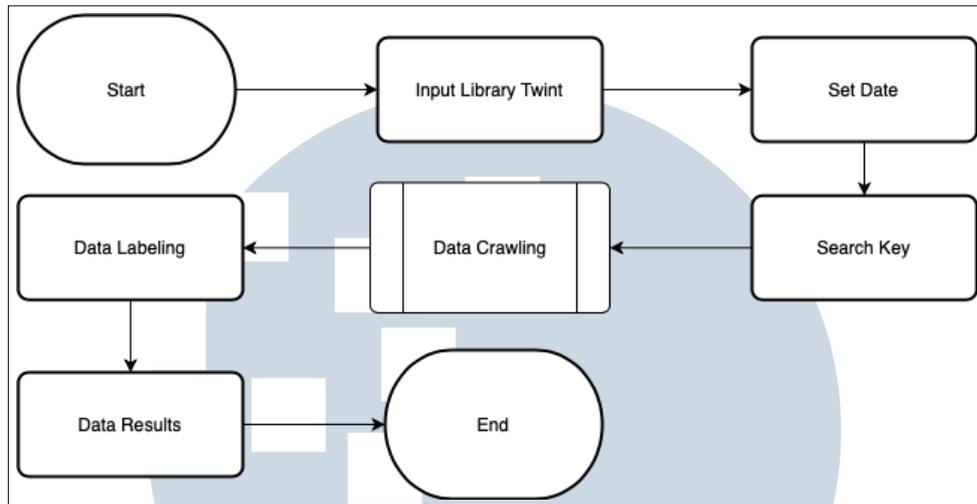
Pada bagian ini dijabarkan langkah-langkah yang hendak dilakukan dalam menyusun dan mengerjakan penelitian. Langkah-langkah penelitian yang dijabarkan dimulai dari awal hingga akhir selesai Beberapa contohnya yang dapat diikuti antara lain pengumpulan data, pengambilan sampel, analisis kebutuhan, perancangan, implementasi, dan pengujian yang dapat dilihat pada Gambar 3.1.



Gambar 3.1. Struktur metodologi penelitian

3.2 Data Crawling

Crawling adalah suatu teknik yang digunakan untuk mengumpulkan informasi yang ada dalam web. Crawling data dilakukan menggunakan *library twint* untuk mendapatkan sejumlah *tweet* pengguna. Gambar 3.2 merupakan alur dalam melakukan crawling data pada waktu yang ditentukan dengan bantuan *twint* yang menggunakan *search key* 'RUU PDP', 'Perlindungan Data Pribadi' sebagai kata kunci dalam mengumpulkan data.[16]



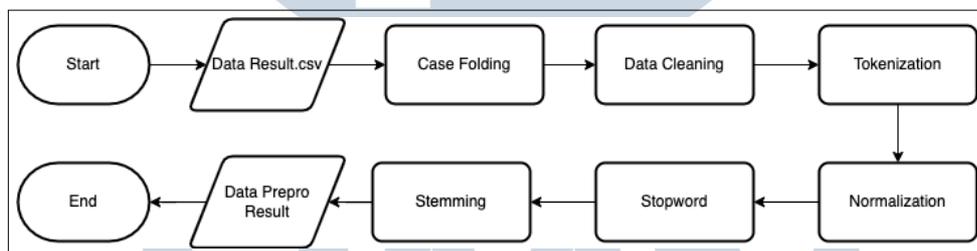
Gambar 3.2. Flowchart data crawling

Sumber data didapatkan dari Twitter, *crawling* akan mengumpulkan data yang kata kuncinya sudah ditentukan yaitu "RUU PDP" dan "Perlindungan data pribadi". Pengambilan tweet dilakukan dengan menggunakan *library twint* yang merupakan *scrapping tools* yang telah disediakan oleh *python*. Setelah mendapatkan data yang diinginkan, data akan disimpan dalam bentuk CSV. Lalu akan dilakukan *labeling data* yang bertujuan untuk memisahkan emosi dari cuitan yang diunggah oleh para pengguna twitter mengenai RUU PDP, *labeling data* yang berisikan tentang opini, saran, atau emosi positif seperti senang dan bahagia. Kelas negatif yang berisikan cuitan kritik, keluhan dan emosi negatif seperti marah dan kecewa. Dan termasuk dalam kelas netral yang berisikan emosi biasa saja atau tidak pada kedua kelas positif dan negatif seperti pertanyaan.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.3 Preprocessing

Preprocessing atau *text processing* adalah proses otomatis menganalisis dan menyortir data teks yang tidak terstruktur untuk mendapatkan informasi yang diinginkan. *Natural language processing* (NLP) dan *machine learning, text processing* dapat secara otomatis memahami bahasa manusia dan mengekstrak nilai dari data teks. Oleh karena itu, penting untuk melakukan pembersihan data sebelum dilakukan klasifikasi, dikarenakan data teks yang didapat mengandung banyak bagian yang tidak terlalu penting seperti *URL link, emoticon* dan sejenisnya. Selain itu, banyak kata - kata yang tidak terlalu berdampak pada orientasi umumnya. Mempertahankan kata - kata tersebut dapat membuat dimensi masalah menjadi tinggi dan menyebabkan klasifikasi menjadi sulit dikarenakan setiap kata dalam teks diperlakukan sebagai satu dimensi. Menghilangkan masalah tersebut pada teks dapat membantu meningkatkan kinerja klasifikasi dan mempercepat prosesnya. Proses pembersihan teks diantaranya adalah : *case folding, data cleaning, tokenizing, normalization, stopword removal* dan *stemming*. [17]



Gambar 3.3. *Flowchart Preprocessing*

Berikut tahapan dalam *text preprocessing* :

1. *Case Folding*: *Case folding* berfungsi untuk melakukan *lower case*, atau mengubah huruf kapital menjadi huruf kecil.
2. *Data Cleaning* : Data mentah yang telah diperoleh perlu diseleksi kembali, seperti menghapus *hashtag*, isi *tweets* yang bukan alfabet (*emoji*), *mention, username, url, dan email* pada dataset yang telah disatukan.
3. *Tokenizing* : Proses memisahkan dari sebuah kalimat menjadi daftar kata tunggal, agar dapat mempermudah proses *stemming* dalam menganalisis untuk mencari kata dasar. Contoh pemisahan kalimat menjadi token sebagai berikut :

saya tidak setuju dengan kebijakan tersebut

Tabel 3.1. Contoh kata sebelum dilakukan tokenizing

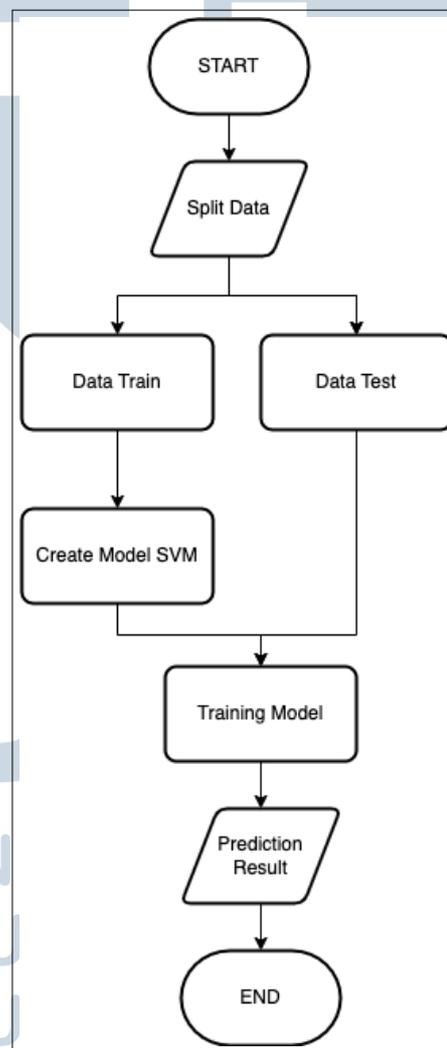
Output : | *saya* | | *tidak* | | *setuju* | | *dengan* | | *kebijakan* | | *tersebut* |

4. *Normalization* : Normalisasi berfungsi untuk mengembalikan kata singkat yang tidak baku menjadi kata yang baku contoh seperti kata "kemn" menjadi kata "kemana".
5. *Stopword Removal* : Merupakan tahap pengambilan kata-kata penting dari hasil token dengan menggunakan algoritma stoplist (menghilangkan kata yang tidak penting).
6. *Stemming* : Merupakan teknik yang digunakan untuk mengekstrak bentuk dasar kata dengan menghilangkan imbuhan darinya. Misalnya, akar kata dari "kemana" adalah "mana". *Stemming* berguna untuk mengindeks kata-kata, dengan cara ini *stemming* mengurangi ukuran indeks dan meningkatkan akurasi pengambilan.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.4 Analysis Sentiment With SVM

Analisis data dilakukan menggunakan algoritma *Support Vector Machine* karena bekerja dengan baik dengan batas pemisahan yang jelas dan dengan ruang berdimensi tinggi. Tetapi sebelum metode di implementasikan, data terlebih dahulu di uji dan di latih. Selanjutnya didapatkan hasil data testing menggunakan metode *Support Vector Machine*, berikut pada Gambar 3.4 menampilkan alur kerja metode *Support Vector Machine* dengan dataset.



Gambar 3.4. *Flowchart Classification SVM*