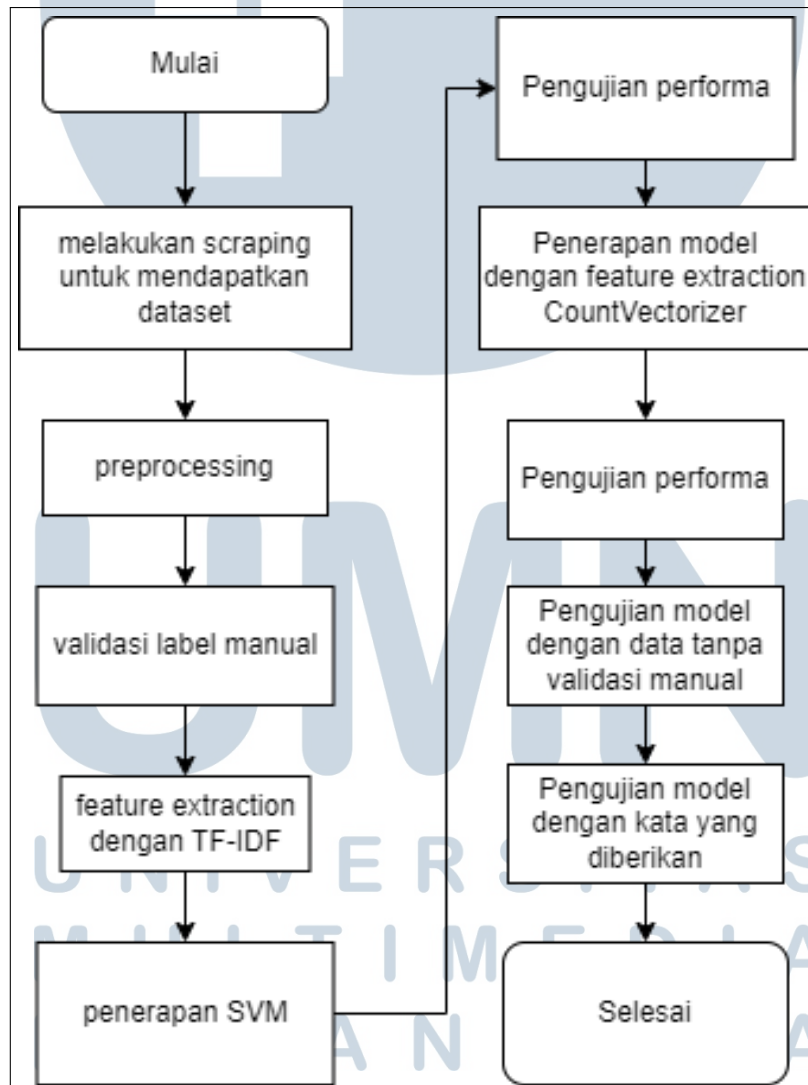


BAB 3 METODOLOGI PENELITIAN

Untuk mencapai tujuan dari yang telah disebutkan, perlunya metodologi sebagai langkah dalam penelitian. Berikut merupakan langkah yang diambil dalam menjalankan penelitian. Literatur yang diambil berfokus pada cara kerja SVM, pengambilan *dataset* dengan model *scraping*, *preprocessing* data dan hal lain yang mendukung dalam penelitian. Proses pada pembuatan model terlampir pada Gambar 3.1 dalam bentuk *flowchart*.



Gambar 3.1. Flowchart proses pembuatan model

3.1 Studi Literatur

Studi literatur dilakukan dengan mencari dan membaca hasil penelitian serta *paper* terdahulu, serta mencari teori yang dibutuhkan dalam penelitian. Pencarian hasil penelitian berfokus pada *feature extraction*, *preprocessing*, pengertian sentimen analisis, SVM, dan cara perhitungan performa.

3.2 Pencarian Dataset

Dataset yang digunakan didapatkan dengan cara melakukan *scraping* pada Google Playstore terkhusus pada aplikasi PeduliLindungi [16]. *Dataset* yang diambil menurut bintang atau ulasan yang diberikan oleh pengguna aplikasi, dan berimbang antara positif dan negatif. Pembagian sentimen negatif dan positif dilakukan secara manual.

3.2.1 Scraping Google Playstore

Scraping Google Playstore menggunakan *library* google-play-scraper. Pada tahap ini, di awal dilakukan inisiasi dengan cara mencoba untuk terhubung dengan *package* PeduliLindungi. Pengambilan menggunakan beberapa tahap, yaitu pengambilan 400 bintang 5, 400 bintang 1, 200 bintang 2, 200 bintang 3, 200 bintang 4. Semua data lalu dijadikan 1 *dataset*.

3.2.2 Manual Labelling

Data yang telah tersedia, dilakukan *labelling* secara manual dengan berdasarkan bintang, jika bintang 5 maka akan dilabeli 'positive', jika 1 akan dilabeli 'negative', dan sisanya akan dilabeli dengan 'neutral'. Pada tahap setelah *preprocessing*, akan dilakukan validasi pada data sehingga data yang telah di label sebelumnya diperiksa kembali apakah label yang diberikan sesuai dengan sentimen yang ada. Sebelum tahap ini dilakukan, dilakukan juga menghilangkan data yang hanya mengandung simbol atau tidak memiliki kata sama sekali, dikarenakan tidak adanya sentimen yang muncul. Validasi ini dilakukan secara manual, yaitu membaca data satu persatu dan memastikan ter label dengan benar. Dalam tahap ini juga dilakukan validasi kembali sehingga mendapatkan masukan terkait validasi yang telah dibuat. Hasil dari validasi manual yang dilakukan juga dicheck kembali sehingga mengurangi unsur subjektif dalam pemberian label.

3.3 Implementasi

Implementasi algoritme dan penggunaan *dataset*, serta training dan testing data. Sebelum pengimplementasian algoritme, *dataset* yang telah ada melalui beberapa tahap *preprocessing*, yaitu tahap menghilangkan simbol, memperkecil kata, dan *stemming* serta *tokenizing* [7]. *Dataset* akan dibagi dalam 80% untuk *training*, dan 20% untuk *testing*. Pengukuran di sini yaitu pada akurasi model dalam membaca 20 persen data dari *testing*.

3.3.1 Preprocessing

Preprocessing terdiri dari beberapa tahap. Tahap pertama, data pada kolom 'content' pada *dataset* akan di buat menjadi tanpa huruf kapital, dengan tujuan dapat mendeteksi persamaan kata tanpa terpengaruh huruf kapital. Proses selanjutnya adalah menghilangkan tanda baca dalam kalimat, sehingga dapat dianalisa per kata tanpa terganggu tanda baca. Dalam tahap ini juga menghilangkan simbol sehingga murni hanya berisikan kata saja. *Stemming* atau menghilangkan imbuhan juga dilakukan, dengan menggunakan library Sastrawi, sehingga didapatkan hanya kata dasar dan menjadi lebih akurat dalam penerapan algoritme. Tahap terakhir dalam *preprocessing* merupakan tokenisasi, yaitu membagi antara 1 kata dengan kata yang lain dalam suatu kalimat.

3.4 Penerapan Algoritme

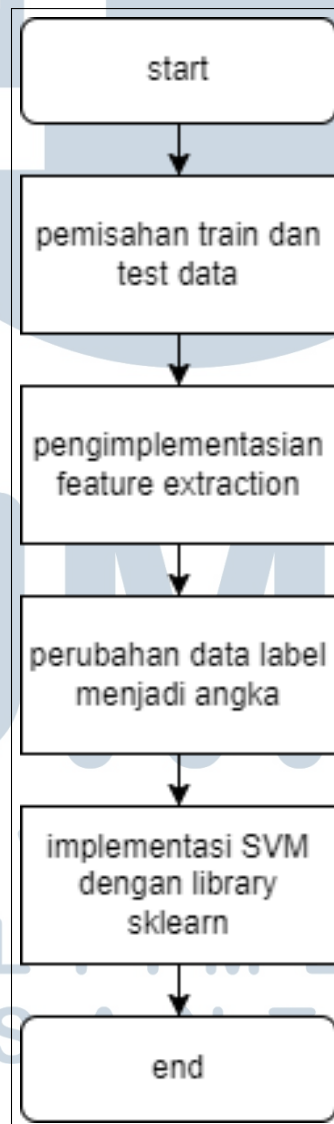
Penerapan algoritma akan menjelaskan bagaimana hasil ulasan yang telah diolah agar dapat dibaca dan dipelajari oleh algoritme sehingga membentuk sebuah model. Penerapan algoritme termasuk di dalamnya terdapat penerapan *feature extraction* dan SVM.

3.4.1 Feature Extraction

Penggunaan TF-IDF dipilih dalam penerapan *feature extraction*. Metode ini juga dibandingkan dengan metode lain yaitu metode CountVectorizer sebagai pembandingan metode TF-IDF untuk melihat tingkat akurasi kedua metode tersebut.

3.4.2 Implementasi Support Vector Machines

Data yang telah diolah dalam proses sebelumnya akan diterapkan dalam proses ini. Pada proses sebelumnya kata telah dilakukan *preprocessing* pada *train_x* dan *test_x* yang pada awalnya berisikan kata dari hasil *preprocessing* yang berisi komentar ulasan PeduliLindungi. Pada *train_y* dan *test_y* berisikan data setimen dari komentar ('positif', 'negative', 'neutral'). Penerapan algoritme menggunakan 20% dari data untuk test dan 80% dari data untuk *training*. Penerapan SVM dilakukan beberapa kali untuk melihat beberapa cara ataupun metode yang paling efektif dalam *preprocessing* dan *feature extraction*. *flowchart*.



Gambar 3.2. Flowchart proses implementasi SVM

3.5 Pengujian

Pengujian yang dilakukan dengan menghitung akurasi, *f1*, *precision*, dan *recall* dari hasil yang didapatkan. Hasil pengujian dilakukan juga untuk melakukan *tuning* kembali pada model, atau membandingkan proses yang telah dilakukan dalam membentuk model seperti validasi dan *labelling* serta *preprocessing*.

3.6 Evaluasi

Evaluasi dari hasil pengujian yang ada. Evaluasi akan dilihat dari akurasi model juga bagaimana model dengan SVM dapat memprediksi sentimen dari *input* yang diberikan, dan melakukan dokumentasi dalam bentuk laporan penelitian.

