



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Objek Penelitian

Objek Penelitian dari topik proposal skripsi ini adalah pendapat ataupun sentimen pada masyarakat Indonesia tentang jenis vaksin COVID-19 yang dimana data yang diambil berasal dari sosial media Twitter. Pendapat yang akan dicari adalah yang bersifat positif, netral dan juga negatif terkait jenis vaksin COVID-19 yang beredar di Indonesia pada Februari 2022 dengan jumlah data sebanyak 4.585 data teks. Dengan data teks yang sudah dikumpulkan tadi akan dilakukan analisa sentiment masyarakat. Lalu, dengan data teks tersebut akan dianalisa dan digunakan untuk mencari tahu sentiment yang ada pada masyarakat dengan menggunakan algoritma klasifikasi.

#### 3.2. Variabel Penelitian

Sesuai dengan penelitian yang akan dilakukan ini, variabel penelitian dari skripsi ini adalah pendapat masyarakat di Indonesia tentang jenis-jenis vaksin COVID-19 pada sosial media Twitter, terdapat dua variabel yaitu dependen dan independent, dimana data independent nantinya akan berisi opini yang ada pada masyarakat Indonesia tentang jenis vaksin COVID-19 di Twitter, dan variable dependennya adalah tanggapan masyarakat terhadap vaksin COVID-19.

#### 3.3. Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan untuk melakukan penelitian ini adalah data teks yang diambil lewat sosial media Twitter dengan bantuan software *RapidMiner Studio*. Data yang diambil adalah tweet yang ada pada bulan Februari

serta April 2022 dengan beberapa kata kunci nama vaksin seperti “Vaksin Sinovac”, “Vaksin AstraZeneca”, “Vaksin Sinopharm”, “Vaksin Moderna”, “Vaksin Pfizer” dan tweet yang diambil hanya yang menggunakan Bahasa Indonesia. Hasil dari pengumpulan data teks tersebut akan disimpan dalam *excel* untuk mempermudah proses akses dan juga melakukan tahapan selanjutnya.

### 3.4. Metode Penelitian

Dalam hal klasifikasi teks, terdapat beberapa algoritma yang cukup sering digunakan untuk kasus analisa sentiment, seperti *Decision Tree* dan juga SVM. Berikut ini adalah perbandingan kedua algoritma tersebut yang akan dijabarkan dalam bentuk tabel[34][35].

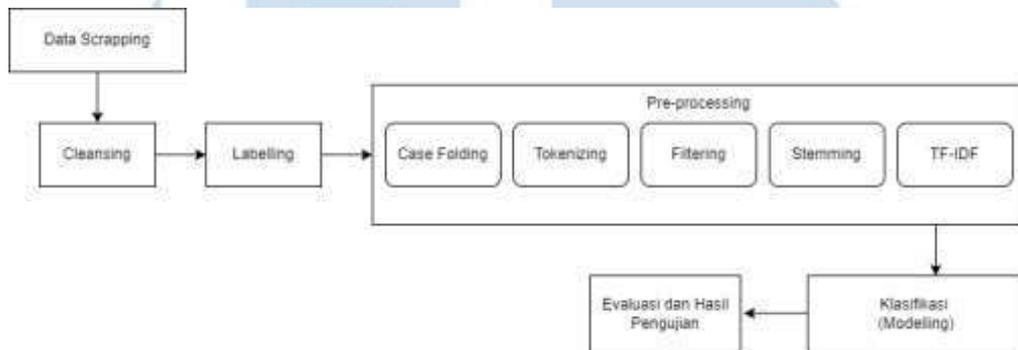
**Tabel 3. 1 Tabel Perbandingan Decision Tree dan SVM**

Decision Tree	Support Vector Machine
<i>Supervised</i> dan <i>unsupervised learning algorithm</i>	<i>Supervised learning algorithm</i>
Digunakan pada prediksi dan klasifikasi tetapi menggunakan 2 faktor.	Digunakan pada umumnya untuk klasifikasi dan regresi.
Bisa digunakan sebagai prediksi maupun klasifikasi.	Bisa digunakan sebagai klasifikasi teks

Berdasarkan Tabel 3.1 perbandingan antara algoritma *Decision Tree* dan juga SVM, algoritma *Decision Tree* dikenal sebagai sebuah *machine learning* yang bisa melakukan prediksi dan klasifikasi. Sedangkan, algoritma SVM dikenal sebagai algoritma yang melakukan klasifikasi teks saja.

### 3.4.1. Kerangka Pikir

Kerangka berpikir yang akan digunakan pada penelitian ini adalah sebagai berikut:



**Gambar 3. 1 Kerangka Pikir Penelitian Modifikasi**

**Sumber:** [36][37]

Pada awal penelitian, langkah yang akan dilakukan adalah data scrapping dengan software *RapidMiner*, pada software *RapidMiner* harus dilakukan pendaftaran hak akses dengan akun *RapidMiner* sehingga proses data scrapping dari *Twitter* bisa dijalankan. Untuk data yang diambil adalah dengan kata kunci nama-nama vaksin yang ada di Indonesia. Setelah itu, data teks yang sudah terkumpul akan melewati proses *cleansing* dan pelabelan. Kemudian, datateks yang sudah dikumpulkan akan melewati serangkaian proses pada *pre-processing* seperti *case folding*, *tokenizing*, *filtering*, *stemming*. Setelah tahap *preprocessing* selesai maka akan dilanjutkan dengan pelatihan model algoritma dengan membagi data train dan data test dengan menentukan jumlah *K-fold Cross Validation* dimana data train akan digunakan untuk membentuk model semula dan data test

akan menguji model yang sudah terbentuk. Lalu, tahapan terakhir adalah evaluasi hasil yang berisikan hasil dari pengujian algoritma *Decision Tree* dan juga SVM.

### 3.4.2. *Text Preprocessing*

Data yang sudah dikumpulkan tadi, akan dilakukan serangkaian proses sehingga bisa mempermudah proses analisis sehingga menjadi data yang siap diolah dengan bantuan *RapidMiner Studio*. Tahap *preprocessing* terdiri beberapa tahapan seperti:

1. *Case folding*: merubah semua huruf menjadi huruf kecil karena kondisi teks yang berbeda.
2. *Tokenize*: pada tahap ini akan dilakukan pemisahan kalimat pada tweet menjadi potongan kata.
3. *Filtering*: memfilter kata-kata yang dianggap kurang penting seperti kata penghubung.
4. *Stemming*: proses ini akan menghilangkan kata imbuhan.
5. TD-IDF: TD-IDF dilakukan untuk memberikan pembobotan pada tiap kata dalam tweets. TD-IDF akan dilakukan langsung dengan *software RapidMiner* karena tersedia fitur *process documents form data*. Tujuan dari pembobotan ini adalah untuk menambah nilai pada algoritma dalam membuat suatu model sehingga membantu algoritma dalam membentuk model yang lebih akurat.

### 3.4.3. *Modelling*

Proses ini membutuhkan pembagian data, dimana data utama akan dibagi menjadi dua jenis yaitu data training dan data testing. Proses pengujian akurasi ini dilakukan dengan metode *K-fold Cross Validation*, yang dimana dengan bantuan metode tersebut bertujuan untuk mengukur hasil akurasi dari model yang sudah dihasilkan dan bisa menemukan hasil yang paling akurat. Pada penelitian ini akan menggunakan *5-fold Cross Validation* karena pada penelitian ini mengadopsi pembagian *data train* dan *data testing* dengan bobot 80:20.

### 3.4.4. *Evaluasi*

Setelah melewati serangkaian proses di atas, maka hasil dari pengujian algoritma tersebut akan ditampilkan dengan menggunakan *confusion matrix* dan hasil dari akurasi algoritma akan ditampilkan, jika semakin tinggi maka akan semakin baik.

