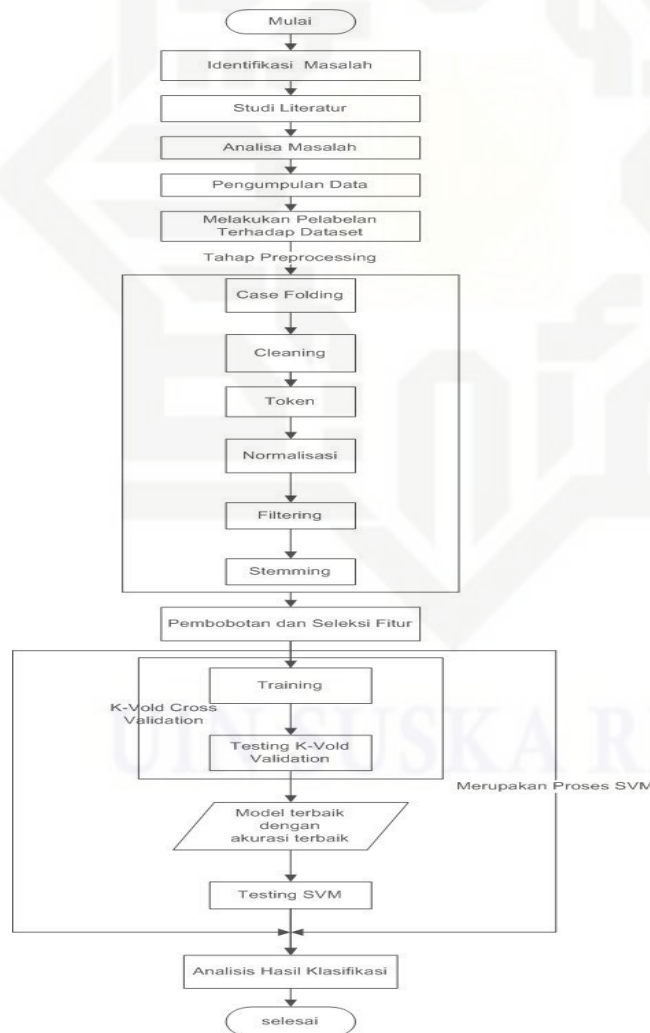


## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Tahapan Penelitian

Metodologi penelitian merupakan salah satu acuan dalam pelaksanaan suatu penelitian. Di dalam metodologi berisikan rencana kerja yang telah terstruktur, agar hasil yang diperoleh sesuai dengan yang diharapkan. Berikut adalah metodologi yang digunakan dalam penelitian tugas akhir ini dengan judul “Klasifikasi Emosi dari Tweet Bahasa Indonesia Menggunakan Metode *Support Vector Machine*” seperti pada gambar berikut ini :



**Gambar 3.1 Tahapan Penelitian**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 3.2 Identifikasi Masalah

Identifikasi masalah merupakan langkah awal dalam melakukan klasifikasi emosi pada *tweet* khususnya *tweet* berbahasa Indonesia, serta berdasarkan hasil dan fakta dari penelitian terkait yang telah dilakukan sebelumnya. Tahap ini juga menerapkan pengenalan Twitter dan *supervised learning*.

## 3.3 Studi Literatur

Acuan dari penelitian yang dilakukan yaitu dengan mencari informasi berdasarkan kajian teoritis yang berhubungan dengan *supervised learning*, *tweet*, *text mining*, SVM dan literatur-literatur yang ada. Literatur yang digunakan adalah jurnal-jurnal penelitian yang terkait dengan materi yang diteliti oleh peneliti sebelumnya untuk membantu dalam pemahaman langkah-langkah klasifikasi jenis emosi pada *tweet* berbahasa Indonesia menggunakan metode SVM.

## 3.4 Analisa Masalah

Analisa merupakan sebuah tahapan yang didalamnya mengevaluasi permasalahan serta mempelajari permasalahan tersebut. Adapun rincian analisa adalah sebagai berikut :

### 3.4.1 Analisa Kebutuhan Data

Menganalisa data masukan yang dikumpulkan serta digunakan dalam penelitian ini. Adapun rincian analisa kebutuhan data yang akan digunakan dalam proses analisa adalah sebagai berikut :

1. Kategori Data

Penelitian ini membutuhkan 5 kategori emosi pada *tweet* yaitu kategori marah, kategori senang, kategori sedih, kategori takut dan kategori jijik.

2. Pengumpulan Data

Data yang digunakan di *download* dari server Twitter dengan memanfaatkan Twitter API (*Application Programming Interface*). *Tweet* yang diambil melalui akun berdasarkan *username* dan pengunduhan jumlah *tweet* dilakukan secara acak per setiap akun Twitter. Kemudian setelah semua data berhasil diunduh dari beberapa akun, data tersebut

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

diseleksi terlebih dahulu untuk mengambil *tweet* yang berkategori marah sebanyak 400 data, kategori senang 400 data, kategori sedih 400 data, kategori takut 400 data dan kategori jijik 400 data. Jadi jumlah keseluruhan data yang digunakan pada penelitian ini adalah sebanyak 2000 *dataset*.

### 3. Pelabelan Data

Setelah mengumpulkan data *tweet* serta menyeleksinya, tahap selanjutnya adalah memberi label pada seluruh data *tweet* sesuai dengan kategori emosi yang telah ditentukan. *Dataset* tersebut kemudian dibagi menjadi 2 yaitu data latih dan data uji. Adapun perbandingan yang digunakan adalah 80% : 20%, (80% sebagai data latih dan 20% sebagai data uji), 70% : 30%, (70% sebagai data latih dan 30% sebagai data uji), dan 90% : 10%, (90% sebagai data latih dan 10% sebagai data uji). Untuk setiap *tweet* diberi label oleh ahli psikolog dan melihat referensi terhadap penelitian terkait sebelumnya.

## 3.5 Analisa Preprocessing

Tahapan *preprocessing* merupakan tahapan awal dari *text mining* yang bertujuan untuk mempersiapkan teks menjadi data yang akan diolah pada tahapan berikutnya. Tahapan *preprocessing* juga merupakan suatu proses dalam melakukan pembersihan kata dari beberapa dokumen yang sebelumnya telah diperoleh dari twitter API. Tujuan dari *text preprocessing* ini adalah untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi *volume* kata. Pada tahapan *preprocessing* ini terdiri dari beberapa proses yaitu :

#### 1. Case Folding

Proses penyeragaman semua huruf menjadi huruf kecil.

#### 2. Cleaning

Proses penghilangan karakter atau tanda baca yang tidak diperlukan dari teks tersebut. Misalnya menghilangkan *mention*, URL, *emoticon*, tanda seru, tanda titik, tanda koma dan tanda tanya.

#### 3. Token

Proses untuk memecah kalimat menjadi beberapa bagian atau kata.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Normalisasi

Proses koreksi ejaan kata terhadap kata yang tidak standar, misalnya “tiduur” menjadi “tidur”. Terhadap singkatan yang tidak baku seperti “tdk” menjadi “tidak”.

5. *Filtering*

Proses untuk menghilangkan kata-kata yang dianggap tidak memiliki makna selama proses klasifikasi. Alasan penghapusan *stopword* dalam banyak program aplikasi yang berkaitan dengan *text mining* adalah karena penggunaannya sangat umum sehingga pengguna dapat berfokus pada kata-kata yang jauh lebih penting.

6. *Stemming*

Penentuan kata dasar dari suatu kata yang memiliki imbuhan, akhiran, awalan, sisipan, awalan dan akhiran. Algoritma yang digunakan adalah ECS (*Enhanced Confix Stripping*).

### 3.6 Pembobotan dan Seleksi Fitur

Dalam tahapan pembobotan dan seleksi fitur ini menggunakan TF-IDF, berfungsi untuk mengurangi *noise*, agar mampu meningkatkan akurasi terhadap klasifikasi. Pemilihan fitur juga berfungsi untuk mendapat nilai *threshold* parameter dalam klasifikasi SVM nantinya. Hal tersebut dilakukan dengan melakukan pengamatan terhadap distribusi frekuensi kemunculan kata dan jumlah *feature*.

### 3.7 Proses SVM

Pada proses SVM, terdapat proses *training* dan *testing*. Proses *training* dengan *dataset* adalah proses melatih mesin, dimana mesin diberi pengetahuan dengan sekumpulan *dataset* yang telah dilabel sebelumnya. Kemudian proses *testing*, mesin akan diuji menggunakan data baru untuk mengetahui tingkat akurasi dan klasifikasinya. Adapun yang dibutuhkan dalam proses SVM ini adalah:

1. Menentukan fungsi kernel yang akan digunakan pada SVM Multikelas. Kernel yang digunakan adalah kernel *Radial Basis Function* (RBF)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sebagai permodelan *hyperplane* dengan menggunakan parameter C dan Gamma.

2. Menentukan jumlah permodelan *k-fold cross validation* untuk menghasilkan model terbaik dengan menggunakan parameter C dan Gamma.
3. Menentukan nilai parameter C dan Gamma yang akan digunakan sebagai permodelan *hyperplane* SVM agar dapat menghasilkan tingkat akurasi yang tinggi pada saat proses pengujian.

### 3.8 Analisa Hasil Klasifikasi

Untuk tahapan terakhir ini adalah melakukan analisa terhadap hasil akhir dalam klasifikasi jenis emosi yang telah dilakukan. Serta memberikan saran yang membangun agar dapat menimbulkan penelitian baru dan terciptanya klasifikasi jenis emosi berdasarkan *tweet* berbahasa Indonesia yang lebih akurat dan lebih baik lagi.