

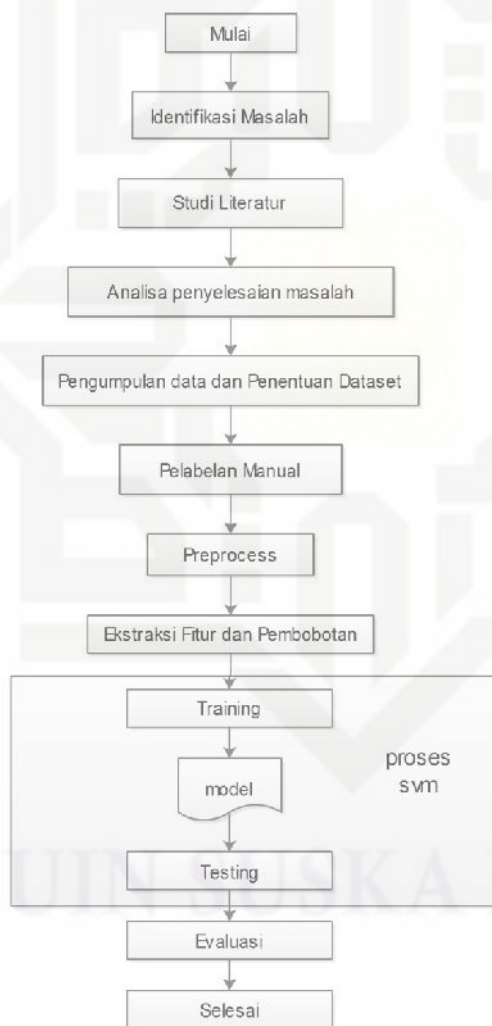
Hak Cipta Dilindungi Undang-undang

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.

BAB III

METODOLOGI PENELITIAN

Metodologi Penelitian adalah sekumpulan peraturan, kegiatan, dan prosedur yang digunakan oleh pelaku suatu disiplin ilmu. Adapun pada penelitian ini, metodologi berupa langkah-langkah yang dilakukan dapat dilihat pada alur berikut:



Gambar VI Metode Penelitian

3.1 Identifikasi Masalah

Pada tahapan ini, identifikasi masalah dilakukan berdasarkan fakta yang ditemukan, setelah melakukan eksplorasi awal. Awalnya, dilakukan eksplorasi untuk menemukan masalah yang terjadi dalam klasifikasi *tweet*. Selain itu eksplorasi juga dilakukan terhadap hal-hal yang dapat digunakan dalam.

3.2 Studi Literatur

Berdasarkan fakta dan temuan masalah yang didapat, maka dilakukan kajian dan studi literatur yang relevan. Teori dan literatur yang dikaji berupa penelitian terdahulu, buku, jurnal dan karya ilmiah, serta situs daring yang berhubungan dengan klasifikasi *tweet*, *e-commerce*.

Pada tahapan ini juga dilakukan kajian terhadap pembelajaran SVM sebagai salah satu metode yang dapat digunakan dalam klasifikasi *tweet*. Hal ini dilakukan untuk mengetahui dan langkah-langkah yang tepat dilakukan untuk mengatasi masalah klasifikasi *tweet e-commerce*.

3.3 Analisa Penyelesaian Masalah

Tahapan ini mulai dilakukan penentuan langkah penyelesaian masalah sesuai dengan batasan penelitian dilakukan. Adapun tahapan yang menjadi analisa terdiri dari pengunduhan dan penentuan *dataset*, pelabelan, *preprocessing*, serta pola pembelajaran SVM.

3.4 Pengumpulan Data dan Penentuan *Dataset*

Pada tahapan ini pengumpulan data dilakukan dengan proses pengunduhan data *tweet* dari internet. Proses-proses pengunduhan dimulai dari proses mendapatkan Twitter API berupa API *key* dan *Access Token*. Dilanjutkan dengan membangun sebuah program aplikasi yang mampu untuk mengunduh entitas *tweet* yang diperlukan untuk *dataset*. Proses tersebut melibatkan *library* *TwitterOauth* yang didapat dari sumber internet.

Setelah data berhasil diunduh data disimpan dalam sebuah *database*. Proses dilanjutkan dengan menyaring seluruh data secara manual untuk menemukan data yang tak dapat menjadi *dataset* pada penelitian ini. Pengunduhan akan dilakukan kembali jika setelah proses penentuan *dataset*, jumlah data yang terdapat pada *database* tidak sesuai dengan jumlah yang ditentukan.

Selain *dataset*, proses pengunduhan juga dilakukan untuk menemukan kamus kata dasar yang digunakan dalam proses penelitian. Kata dasar didapat dari Kamus Besar Bahasa Indonesia versi 4 yang diunduh dari internet.

3.5 Pelabelan Manual

Pelabelan manual dilakukan terhadap *dataset* yang telah tersimpan di *database*. Setiap data dilabel dengan melihat ciri informasi data tersebut secara manual. Label yang telah ditentukan pada penelitian ini berdasarkan teori dan penelitian sebelumnya, yaitu relevan dan iklan.

3.6 Preprocessing

Tahapan ini merupakan persiapan *dataset* menjadi sebuah token tunggal yang berpotensi menjadi sebuah fitur. Pada tahapan ini proses yang dilakukan antara lain :

1. *Case Foding*, proses ini melakukan penyeragaman kata dalam data *tweet*/dokumen
2. *Cleansing*, karakter dan entitas yang tidak diperlukan kemudian dibersihkan dalam proses ini
3. *Tokenizing*, proses ini dilakukan untuk membuat kata-kata dalam dokumen dapat berdiri sendiri menjadi *token*
4. *Convert Word*, proses ini mengubah kata-kata tak baku menjadi kata baku. Hal ini diperlukan agar data *tweet* dapat diproses maksimal pada proses selanjutnya.
5. *Stopword Removal*, proses ini dilakukan untuk menghapus *noise* berupa kata-kata yang tidak diperlukan. Selain menghapus kata-kata *stoplist*, tahapan ini juga menghapus kata dua karakter
6. *Stemming*, proses ini mengubah semua kata menjadi bentuk kata dasar. Hal ini diperlukan untuk mengurangi varietas kata yang diperlukan dalam proses mendapatkan fitur

3.7 Ekstraksi Fitur dan Pembobotan

Tahapan selanjutnya adalah proses ekstraksi fitur dari hasil *preprocessing*. Fitur yang digunakan dalam proses selanjutnya ditentukan dengan menentukan nilai batas bawah (*threshold*) dari bobot keseluruhan *token* hasil *preprocessing*. Penentuan *threshold* dilakukan dengan membuat grafik distribusi fitur untuk melihat nilai konstan jumlah fitur yang dihasilkan dari beberapa rentang nilai.

Selain itu, pada penelitian juga dilakukan proses perangkingan fitur dengan menggunakan bobot dari metode DF untuk menentukan nilai fitur pada proses perangkingan. Pada tahapan ini juga dilakukan pembobotan TFIDF agar hasil *preprocessing* dapat dijadikan *input* dalam proses SVM.

3.8 Proses SVM

Tahapan ini terdiri dari proses *training* dan proses *testing*. Proses ini merupakan pembentukan model pembelajaran klasifikasi beserta hasil uji akurasi. Didalam proses ini juga dilakukan metode pemodelan *k-fold cross validation* dan penemuan parameter terbaik dengan metode *grid search*.

Proses dimulai dari *training* yang menggunakan inputan berupa data *train* dari data *training*. Data tersebut kemudian digunakan sebagai bahan pembelajaran untuk membentuk model yang akan diuji dengan data *testing* untuk mendapatkan akurasi klasifikasi

3.9 Evaluasi Hasil Klasifikasi

Merupakan tahapan pengujian akurasi hasil dari pembelajaran yang dihasilkan dari proses pembelajaran. Tahapan ini merupakan pengukuran akurasi klasifikasi dari model pembelajaran yang diuji.