

BAB III

METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM

3.1 Metodologi Penelitian

Penelitian “Implementasi Algoritma Random Forest Menggunakan TF-IDF Untuk Analisis Sentimen Dengan Penerapan *Transfer Learning*” menggunakan beberapa tahap, tahap-tahap yang dilaksanakan antara lain adalah sebagai berikut.

1. Studi Literatur

Tahap studi literatur dilakukan dengan mencari, membaca dan mempelajari sumber dari jurnal ilmiah dan karya tulis terkait dengan topik yang diteliti. Tahap studi literatur bertujuan untuk memahami teori mengenai TF-IDF, Random Forest Classifier, dan n-grams.

2. Pengumpulan Data

Dataset didapatkan dari ... Dataset yang didapatkan merupakan data ulasan pengguna kepada perusahaan Amazon dan Yelp. Dataset yang digunakan terdiri dari 10000 baris data berupa 5000 berlabel positif dan 5000 berlabel negatif.

3. Perancangan dan Pengembangan Program

Tahap perancangan program terdiri atas pembuatan alur kerja untuk penerapan TF-IDF, Random Forest Classifier, *transfer learning*, dan alur antar muka pada aplikasi web.

4. *Testing* dan *debugging*

Tahap ini bertujuan untuk memastikan dan menguji apakah program *machine learning* dan web yang telah dibangun telah berjalan dengan baik serta melakukan perbaikan jika terdapat kesalahan. Pada program *machine learning* dilakukan pengujian pada hasil dari pembobotan TF-IDF, hasil dari tahap *preprocessing*, dan uji performa dari Random Forest Classifier. Pada program web, dilakukan pengujian apakah website dapat menerapkan model *machine learning* yang telah dikembangkan.

5. Penulisan Laporan.

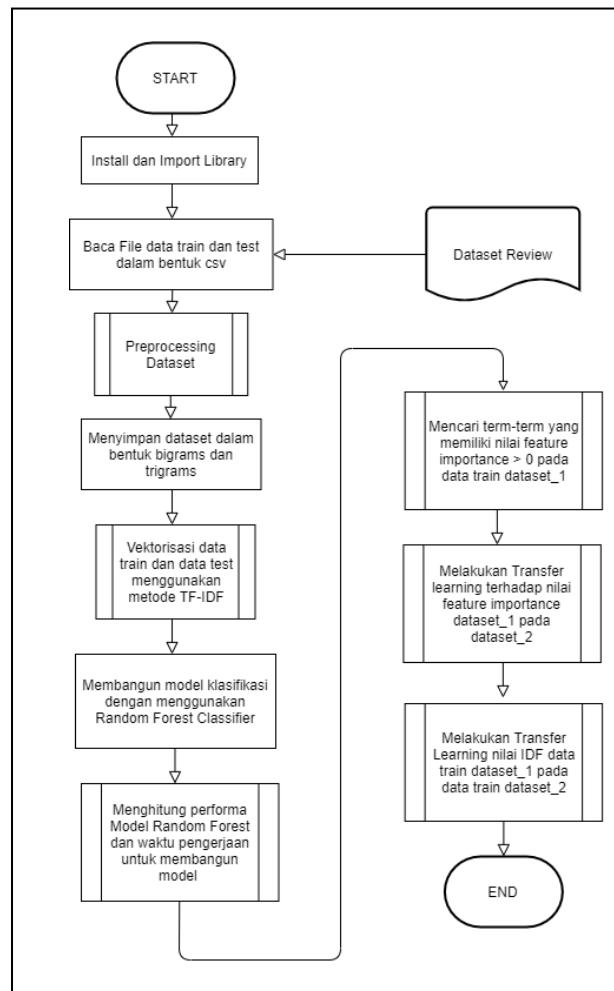
Laporan yang disusun bertujuan sebagai dokumentasi atas penelitian yang telah dilakukan. Penyusunan laporan dibuat secara bertahap, dimulai dari pendahuluan, latar belakang, kesimpulan, dan saran.

3.2 Perancangan Aplikasi

Perancangan aplikasi dilakukan dengan merancang *flowchart* yang menjabarkan tahapan dalam penerapan Random Forest Classifier dan metode *transfer learning* dimulai dari *preprocessing* sampai dengan pembuatan model, serta menjabarkan tentang tahapan dalam penggunaan Random Forest Classifier untuk analisis sentiment dalam aplikasi web. Pada tahap ini juga dilakukan perancangan antarmuka aplikasi web.

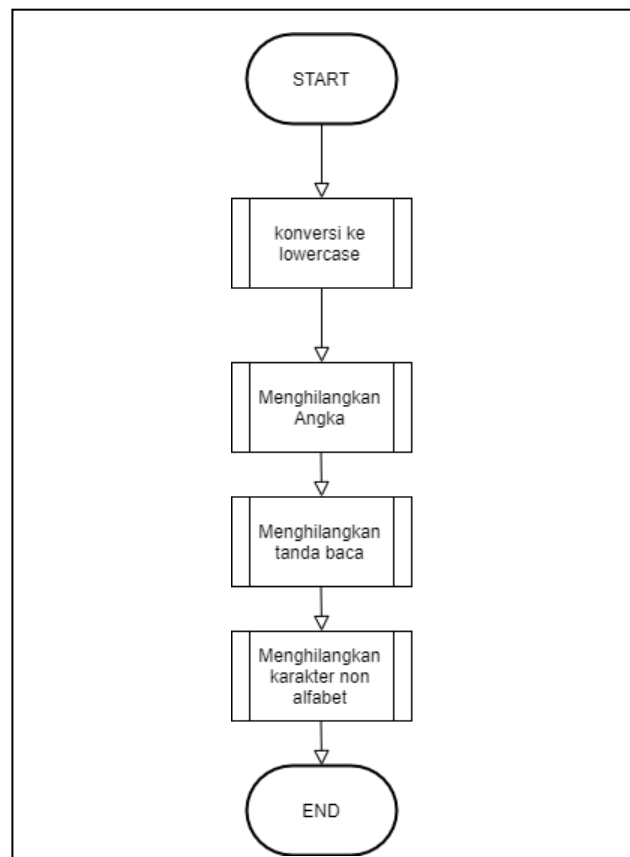
3.2.1 Flowchart

Gambar 3.1 adalah *flowchart* atau diagram alir dari proses klasifikasi sentimen menggunakan TF-IDF dan Random Forest Classifier dengan metode *transfer learning* yang digambarkan secara umum. Proses pertama yang dilakukan adalah *import library* yang dibutuhkan untuk mendukung proses penelitian. Lalu *dataset* dalam bentuk file *csv* yang telah diolah sebelumnya akan di *import* yang selanjutnya akan diproses. *Dataset* yang digunakan adalah *dataset* sentiment ulasan pengguna kepada perusahaan Amazon, Yelp, dan IMDB.



Gambar 3.1 Flowchart utama

Dataset yang telah dibaca terdiri dari data *train* dan data *test* dari masing masing data ulasan tiap perusahaan. Lalu tahap selanjutnya adalah melakukan *preprocessing* pada *dataset*. Dalam proses *preprocessing*, ada beberapa tahap yang dilakukan seperti yang dapat dilihat pada gambar 3.2 Tahap-tahap yang dilakukan antara lain *lowering case* (mengecilkan huruf), menghilangkan angka, manghilangkan *punctuation* (tanda baca), dan menghilangkan karakter yang bukan huruf alfabet. *Preprocessing* dilakukan dengan bantuan *library* string dan re (*regular expression*).



Gambar 3.2 Flowchart tahap *preprocessing*