

BAB 3 METODOLOGI PENELITIAN

Penelitian ini dilakukan secara bertahap. Berikut adalah tahap-tahap yang dilakukan dalam pengerjaan penelitian ini.

3.1 Alur Penelitian

Pada gambar 3.1 merupakan *flowchart* tahap-tahap alur penelitian ini.



Gambar 3.1. *Flowchart* alur penelitian

3.1.1 Studi Literatur

Studi literatur dilakukan dengan cara mengumpulkan informasi yang berkaitan dan mendukung penelitian ini seperti analisis sentimen, *support vector machine*, *text preprocessing*, TF-IDF, SMOTE, dan *Confusion Matrix*. Informasi

ini dikumpulkan berdasarkan sumber literatur yang relevan, seperti jurnal ilmiah, tesis, dan *website* resmi.

3.1.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data terkait penelitian, yang akan digunakan untuk mengembangkan model. Pengumpulan data ini didapatkan dari ulasan yang terdapat di *google play store* menggunakan *library google play scraper*. Kemudian data tersebut akan disimpan dalam bentuk *csv*. Data yang diambil merupakan ulasan sejak Januari 2023 sampai April 2024, karena pada tahun tersebut *shopee* merupakan aplikasi *e-commerce* dengan jumlah pengunjung terbanyak di Indonesia.

3.1.3 Perancangan Sistem

Pada tahap ini akan membahas mengenai perancangan sistem yang akan dibuat. Perancangan sistem ini akan dibuat dalam bentuk *flowchart*. Perancangan sistem ini mencakup beberapa tahapan yang terdiri dari *scraping* data, *text preprocessing*, proses labeling, *split* data, penerapan TF-IDF, penerapan SMOTE, penerapan SVM, dan evaluasi.

3.1.4 Implementasi

Pada tahap ini akan dilakukan implementasi sesuai dengan rancangan sistem yang sudah dibuat. Implementasi ini dilakukan dengan menggunakan bahasa pemrograman *python* yang dijalankan di *jupyter notebook*.

3.1.5 Pengujian Model dan Evaluasi

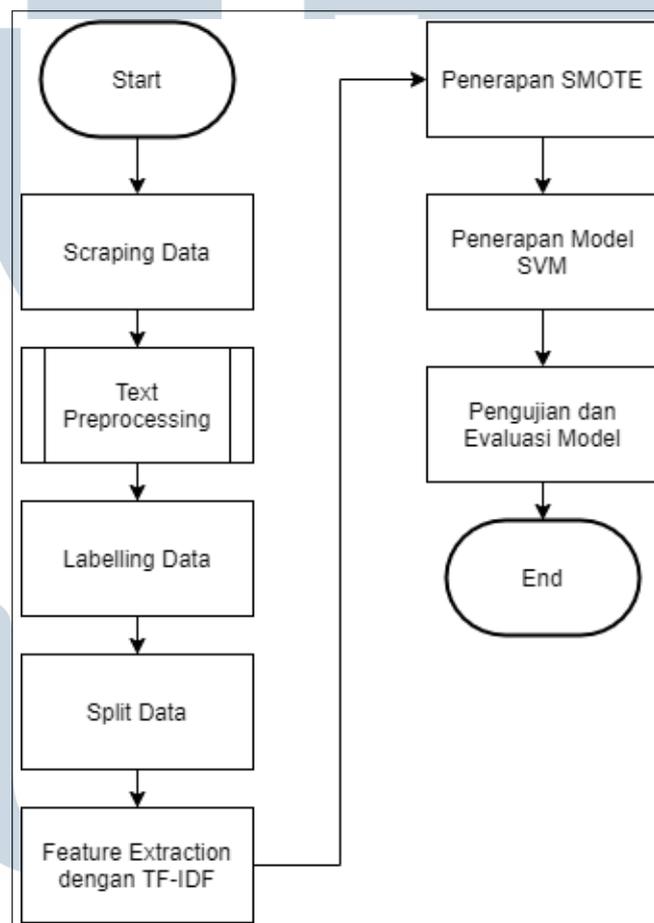
Setelah model berhasil dibuat, akan dilakukan pengujian model menggunakan data *test* yang sudah disiapkan. Setelah itu, model akan dilakukan evaluasi untuk mengukur performa dari model tersebut menggunakan *confusion matrix*.

3.1.6 Pembuatan Laporan

Pada tahap ini akan dilakukan pembuatan laporan untuk mencatat setiap proses yang dilakukan pada penelitian ini.

3.2 Perancangan Sistem

Alur perancangan sistem pada penelitian ini dapat dilihat pada gambar 3.2



Gambar 3.2. Flowchart perancangan sistem

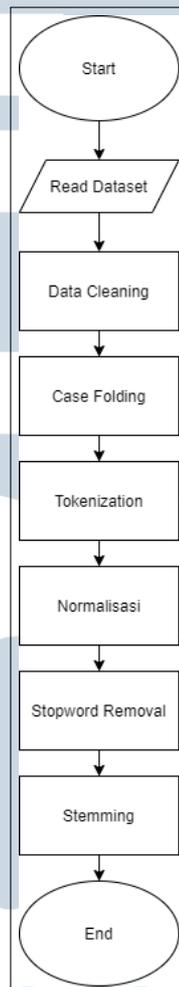
3.2.1 Scraping Data

Pada tahap *scraping* data, akan dilakukan *import library* yang akan digunakan untuk melakukan pengambilan data. *Library* yang digunakan adalah *google play scraper*. Setelah itu, melakukan set *link* aplikasi, bahasa, negara, *sort*,

dan jumlah data yang ingin diambil. Kemudian data akan disimpan dalam bentuk csv.

3.2.2 Text Preprocessing

Pada tahap ini data yang sudah dikumpulkan akan diolah dan dibersihkan. Tahapan *text preprocessing* dapat terlihat pada gambar 3.3.



Gambar 3.3. Flowchart Text Preprocessing

Tahap awal untuk melakukan *text preprocessing* adalah melakukan *import dataset*. Setelah itu, data akan difilter dengan kata kunci "live" yang tidak diikuti dengan kata "chat" sebelum dan sesudah kata "live" untuk mengambil ulasan tentang shopee live. Kemudian data akan melewati beberapa proses, antara lain:

1. *Data Cleaning*

Pada proses ini, data akan dilakukan penghilangan karakter selain huruf dan melakukan penghilangan spasi yang berlebihan. Contoh dari proses ini adalah sebagai berikut: "Aplikasi ini, sangat bagus!" menjadi "Aplikasi ini sangat bagus".

2. *Case Folding*

Pada proses ini, akan dilakukan proses untuk mengubah semua huruf besar pada kalimat menjadi huruf kecil. Sehingga tidak ada lagi kata yang memiliki huruf kapital. Contoh dari proses ini adalah sebagai berikut: "Aplikasi ini sangat bagus" menjadi "aplikasi ini sangat bagus".

3. *Tokenization*

Pada proses ini, akan dilakukan proses untuk memecah kalimat menjadi satuan kata. Contoh dari proses ini adalah sebagai berikut: "aplikasi ini sangat bagus" menjadi ['aplikasi', 'ini', 'sangat', 'bagus'].

4. Normalisasi

Pada proses ini, akan dilakukan proses untuk mengubah kata gaul, singkatan, dan kata tidak baku menjadi kata dasar yang baku. Contoh dari kata yang dinormalisasi adalah "yg" menjadi "yang".

5. *Stopword Removal*

Pada proses ini, akan dilakukan proses untuk penghilangan kata-kata umum yang dianggap tidak penting dan tidak memiliki arti. Contoh dari kata *stopword* adalah "dan", "dari, dan "di".

6. *Stemming*

Pada proses ini, akan dilakukan proses untuk mengubah kata berimbuhan menjadi bentuk kata dasarnya. Contoh dari kata yang dilakukan *stemming* adalah "diskonnya" menjadi "diskon".

3.2.3 Labelling Data

Pada tahap ini, data akan diberikan label secara otomatis menggunakan *library vadersentiment* untuk menentukan sentimen positif, negatif, dan netral. Pertama, data akan diterjemahkan kedalam bahasa inggris menggunakan *library deeptranstralor*. Kemudian data akan diberikan nilai *compound* sesuai dengan

perhitungan yang dilakukan *library vader sentiment*. Jika *nilai* ≥ 0.05 , maka ulasan akan diberikan label positive. Jika nilai *nilai* ≤ -0.05 , maka ulasan akan diberikan label negative. Dan jika *nilai* > -0.05 , dan *nilai* < 0.05 , maka ulasan akan diberikan label neutral.

3.2.4 Split Data

Setelah dilakukan *labelling* data, selanjutnya akan dilakukan pemisahan data menjadi dua bagian, yaitu data *test* dan data *train*. Data *train* akan digunakan untuk melatih algoritma dalam pembuatan model, sedangkan data *test* akan digunakan untuk menguji performa model yang sudah dibuat.

3.2.5 Penerapan TF-IDF

Setelah tahap *split data*, selanjutnya akan dilakukan penerapan TF-IDF untuk menghitung bobot dari setiap kata yang ada di dokumen. Perhitungan ini dilakukan dengan menghitung nilai TF dan nilai IDF pada dataset yang sudah dilakukan *text preprocessing* dan *labeling*. Setelah itu akan dilakukan perhitungan untuk nilai TF-IDF.

3.2.6 Penerapan SMOTE

Pada tahap ini akan dilakukan penyeimbangan data pada kelas minoritas menggunakan SMOTE. Penerapan SMOTE dilakukan setelah data melewati tahap *split* dan TF-IDF. Tahap ini dilakukan dengan tujuan untuk menambahkan jumlah data pada kelas minoritas sehingga data memiliki jumlah yang seimbang dengan jumlah data pada kelas mayoritas.

3.2.7 Penerapan Model dengan SVM

Setelah melalui tahap pembobotan dengan *tfidf*, tahap selanjutnya akan dilakukan pembentukan model dengan algoritma *support vector machine* untuk mengklasifikasikan data. Model akan dilatih dengan data *train* yang sudah melewati tahap *preprocessing* dan *tfidf*.

3.2.8 Pengujian dan Evaluasi Model

Setelah model dilatih menggunakan data *train*, selanjutnya akan dilakukan pengujian dan evaluasi terhadap kinerja model. Model akan dilakukan pengujian menggunakan data *test*. Setelah itu akan dievaluasi menggunakan *confusion matrix*. Hasil dari *confusion matrix* akan digunakan untuk mengukur nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

