

BAB III

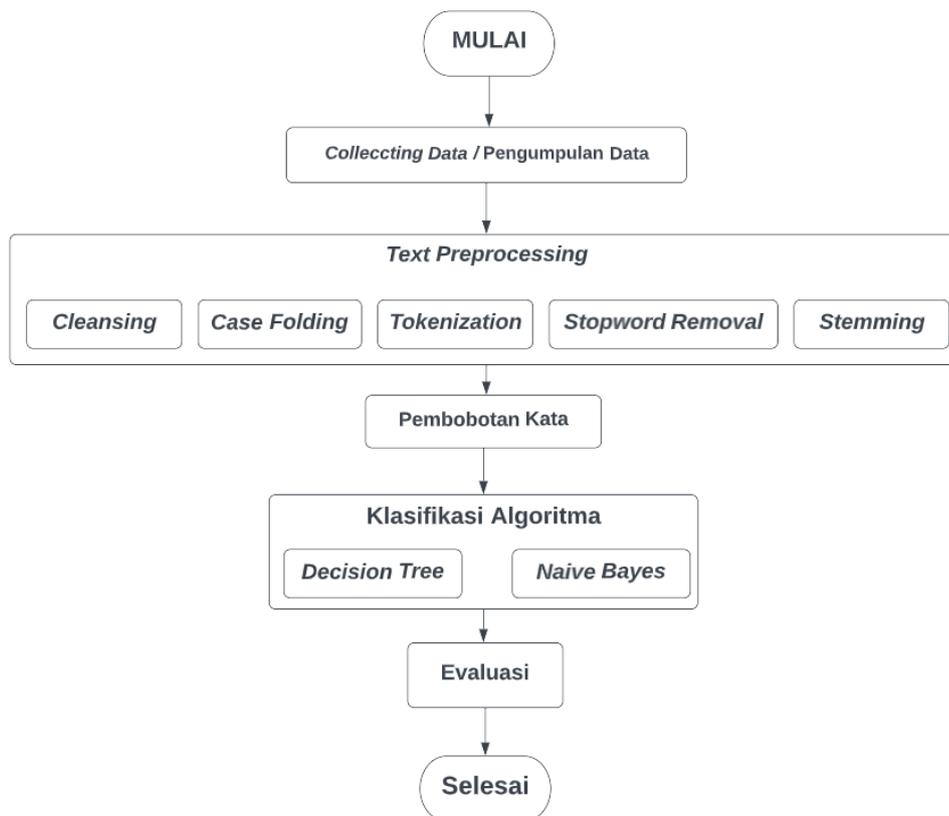
METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek penelitian berisikan tentang gambaran objek yang ada dalam suatu penelitian. Objek penelitian dalam penelitian ini adalah komparasi tingkat akurasi pada algoritma *Decision Tree* dan algoritma *Naive Bayes* dalam analisis sentimen terhadap Tokopedia pada media sosial *Twitter*.

3.2 Metode Penelitian

Metode penelitian pada penelitian berisikan tentang tahapan-tahapan penelitian seperti pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.2.1 Pengolahan Data

Pada tahap pengolahan data dilakukan melalui beberapa proses, yaitu:

1. *Collecting Data*

Collecting data atau disebut juga pengumpulan data ini merupakan proses yang dilakukan dengan melakukan proses *crawling* data dari media sosial *Twitter* dengan menggunakan *software Rapidminer Studio*. Data yang didapatkan berupa *tweet* dari persepsi pengguna *marketplace* Tokopedia. Hasil dari pengumpulan data ini kemudian disimpan ke dalam format *excel (.xlsx)* untuk mempermudah akses dalam melihat hasil dan proses *labelling* data nantinya.

2. *Text Preprocessing*

Text preprocessing adalah proses mengubah data mentah menjadi data yang sesuai dengan prosedur *mining* yang akan dilakukan dan merupakan tahap yang paling penting dalam *data mining* serta berguna untuk menyeleksi data dan mengubahnya menjadi data yang terstruktur. Tahap ini merupakan tahapan awal untuk mengubah struktur isi dari data agar menjadi format yang sesuai dan diproses oleh algoritma *Decision Tree* dan *Naïve Bayes*. *Text preprocessing* ini dilakukan melalui beberapa tahap yaitu:

1) *Cleansing*

Pada tahapan akan dilakukan pembersihan data *tweet* dengan menghilangkan angka, tanda baca, *URL*, nama pengguna, tagar (#), dan *emoticon*. Hasil proses *cleansing* adalah menghilangkan kata dari karakter yang tidak ada artinya dan akan diganti oleh karakter spasi.

2) *Case Folding*

Case folding merupakan proses yang dilakukan untuk mengubah setiap kata yang ada di dalam *dataset* akan di ubah menjadi huruf kecil menggunakan fungsi *lowercase*.

3) *Tokenization*

Pada tahap ini dilakukan pemisahan deretan kata di dalam kalimat, paragraf atau halaman menjadi *token* atau potongan kata tunggal. Pada saat bersamaan, *tokenization* juga membuang karakter selain huruf seperti tanda baca.

4) *Stopword Removal*

Pada tahapan ini kata yang tidak memiliki arti dan tidak berhubungan dengan sentimen akan dihilangkan. Kata dari data *tweet* akan dibandingkan dengan kata yang terdapat dalam *database stopwords*. Hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *database stopwords*.

5) *Stemming*

Pada tahap ini kata akan mengubah kata yang berimbuhan menjadi kata dasar agar memudahkan dalam proses pembobotan.

3. Pembobotan kata

Pada pembobotan kata ini dilakukan proses untuk pemberian nilai atau bobot terhadap setiap kata yang ada pada setiap *tweet* yang telah melewati tahap *preprocessing*. Metode yang digunakan dalam pemberian bobot pada kata yaitu TF-IDF. Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah kata

dalam sebuah *tweet*, dimana nilai dari kata tersebut akan dijadikan sebagai *input* pada proses klasifikasi algoritma nanti.

4. Klasifikasi algoritma *Decision Tree* dan *Naïve Bayes*

Proses selanjutnya melakukan klasifikasi dengan algoritma *Decision Tree* dan *Naïve Bayes*. Pada proses ini dibutuhkan bobot setiap kata yang ada pada data *training*. Nilai bobot setiap kata ini didapat dari proses pembobotan TF-IDF yang akan digunakan dalam pembuatan model dari masing-masing algoritma yang digunakan.

3.2.2 Evaluasi

Pada tahap evaluasi dilakukan pengukuran dan peninjauan hasil penelitian. Pengukuran dan peninjauan bertujuan untuk membandingkan tujuan dari penelitian dengan tingkat keberhasilan yang dicapai melalui pengujian kinerja dari algoritma *Decision Tree* dan *Naïve Bayes* dengan metode *confusion matrix*. *Confusion matrix* merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil yang sebenarnya berdasarkan data *testing* yang akan diproses oleh model klasifikasi *Decision Tree* dan *Naïve Bayes* yang telah dibuat sebelumnya. Metode *confusion matrix* menghasilkan empat nilai yaitu *True Positive*, *True Negative*, *False Positive* dan *False Negative* selanjutnya dilakukan perhitungan mulai dari *accuracy*, *precision*, *recall* dan *f-measure*. Kemudian tahap selanjutnya evaluasi akan menghasilkan informasi yang dapat digunakan untuk penarikan kesimpulan penelitian.