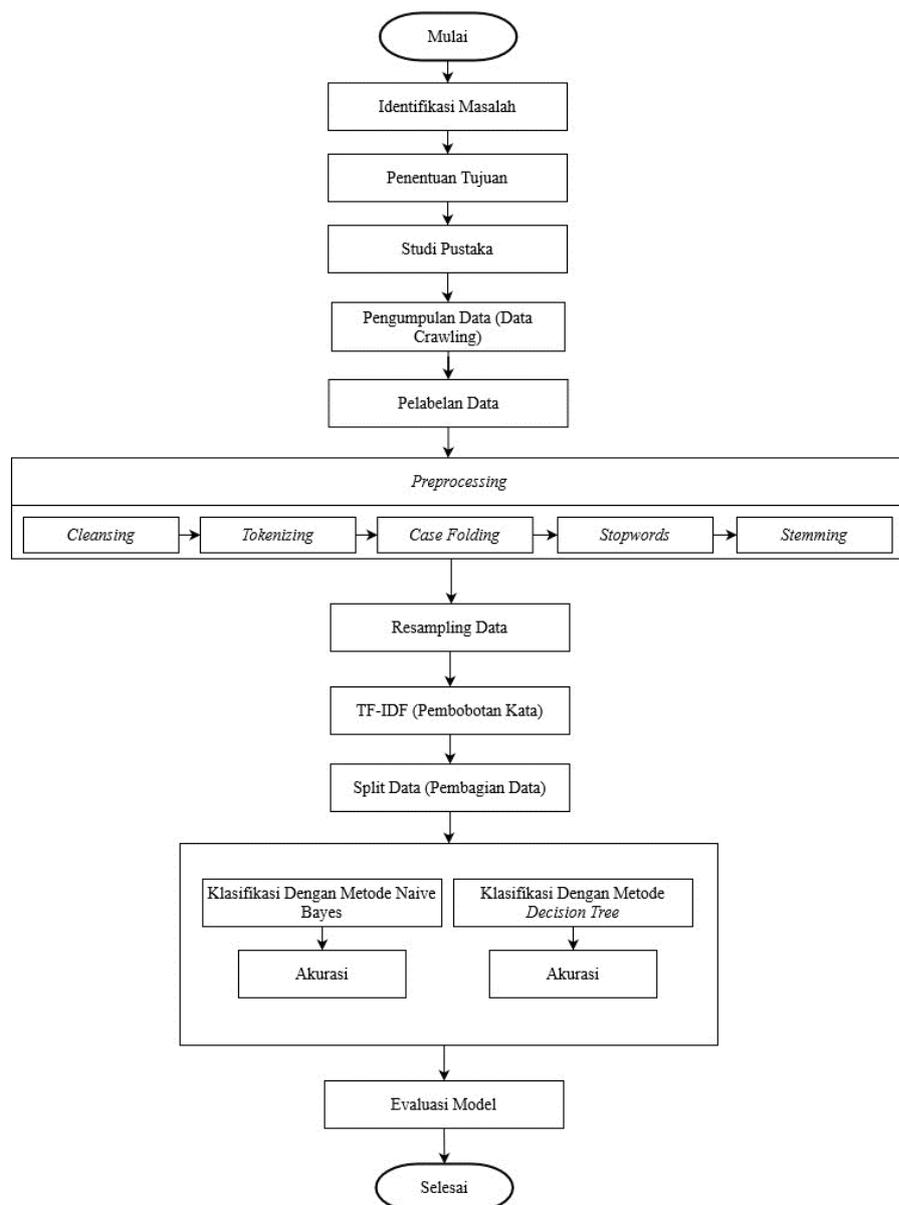


BAB III

METODOLOGI PENELITIAN

Proses dalam melakukan penelitian ini digambarkan dalam sebuah alur pada Gambar 3.1.



Gambar 3.1 Tahapan-Tahapan Penelitian

3.1. Identifikasi Masalah

Mengamati salah satu marketplace yang ada di Indonesia yaitu Tiktok Shop. Banyak opini yang disampaikan oleh masyarakat sehingga bisa dilakukan analisis sentimen pengguna media sosial twitter terhadap Tiktok Shop.

3.2. Penentuan Tujuan

Tahap penentuan tujuan memiliki kegunaan untuk memperjelas sasaran dalam penelitian. Tujuan dari penelitian ini yaitu untuk klasifikasi opini masyarakat yang ada di media sosial twitter tentang Tiktok Shop menggunakan metode *Naïve Bayes* dan *Decision Tree*.

3.3. Studi Pustaka

Studi Pustaka merupakan tahap yang sangat penting dalam proses penelitian. Studi Pustaka diartikan sebagai tahap dimana terjadinya pengumpulan informasi yang berkaitan dengan penelitian yang dilakukan. Hal ini dilakukan dengan mencari bahan-bahan yang menjadi sumber referensi untuk penelitian agar penelitian bisa berjalan dengan baik. Adapun sumber-sumber yang dijadikan referensi yaitu jurnal, tesis, skripsi, buku serta sumber lain yang berkaitan dengan penelitian.

3.4. Pengumpulan Data

Tahap pengumpulan data atau crawling data dibagi menjadi dua yaitu dan kategori data dan pengumpulan data dari media sosial twitter. Penentuan kategori

data akan mempermudah dalam melakukan pengumpulan data. Berikut ini penjelasan dari tahap pengumpulan data.

1. Kategori Data

Kategori atau label data yang digunakan dalam penelitian ini ada tiga yaitu label positif, negatif dan juga netral (Efendi, 2018):

- a. Kategori Positif. Pada kategori ini, data tweet mengandung arti positif tentang Tiktok Shop. Contohnya seperti, “Senang belanja di tiktok shop karena pengiriman yang tidak lama”.
- b. Kategori Negatif. Pada kategori ini, data tweet mengandung arti negatif terhadap Tiktok Shop. Contohnya seperti, “Kenapa barang belum juga sampai? Padahal sudah satu minggu”.
- c. Kategori Netral. Pada kategori ini, data tweet tidak memiliki arti positif atau negatif. Contohnya seperti, “Pendidikan menjadi hal penting #tiktokshop”.

2. Pengumpulan Data

Pada penelitian ini pengumpulan data tweet tentang Tiktok Shop pada media sosial *Twitter* dilakukan dengan metode *crawling*. Dalam teknik ini, Pengambilan data akan diambil dengan menggunakan keyword yaitu #Tiktokshop. Jumlah data yaitu 1000 kemudian akan bagi menjadi data *training* dan data *testing* Instrumen penelitian yang digunakan untuk melakukan eksperimen ini sebagai berikut:

- a. Sistem operasi yang digunakan adalah Windows 10 Pro 64-bit.
- b. Bahasa Pemrograman Python.
- c. Google Colab
- d. Microsoft Excel 2010.

3.5. Pelabelan Data

Pelabelan data adalah tahapan dalam pembelajaran mesin yang bertujuan untuk mengidentifikasi objek dalam data mentah (seperti gambar, video, audio, atau teks) dan menandainya dengan label yang membantu model pembelajaran mesin membuat prediksi dan estimasi yang akurat. Pelabelan data pada penelitian ini dilakukan secara manual dengan tiga kategori data yaitu positif, negatif dan netral.

3.6. Pre-Processing Data

Pada tahap ini, peneliti melakukan proses pengolahan data yang diperoleh dari tahap pengambilan data dengan tujuan untuk menghilangkan beberapa permasalahan yang dapat mengganggu Ketika pemrosesan data.

1. *Cleansing* merupakan teknik yang akan membuat dokumen atau text menjadi bersih dari karakter yang tidak diperlukan seperti *hashtag*, *username*, *url*, *emoticons* dan *HTML*. Pembersihan ini berguna agar noise bisa berkurang sehingga proses klasifikasi bisa berjalan dengan baik.
2. *Tokenizing* merupakan proses dimana dokumen yang tersusun dari kata-kata akan dipecah menjadi kata-kata yang akan memutus urutan string.
3. *Case Folding* merupakan proses pada text preprocessing yang mana pada proses ini dokumen atau text yang terdiri dari susunan huruf akan dirubah menjadi huruf kecil secara keseluruhan supaya tidak terjadi ketidaksamaan di dalam dokumen.

4. *Stopwords*, yakni membuang kosakata yang tidak memiliki arti khusus atau kata-kata unik yang bisa menjadi perwakilan dari dokumen. Contoh kosakata tersebut adalah dan, ini, itu, adalah dan lain sebagainya.
5. *Stemming* merupakan teknik yang berguna untuk mengembalikan kata ke kata dasar karena menghilangkan kata awalan, sisipan, kombinasi dan juga akhiran. Contohnya seperti kata mengambil maka dalam teknik *stemming* akan berubah menjadi ambil.

3.7. Resampling Data

Teknik resampling data yang digunakan pada penelitian ini yaitu *oversampling*. Teknik ini merupakan menambah sample ke kelas minoritas sehingga rasionya sama. Tujuan dari *resampling data* adalah untuk meningkatkan hasil akurasi model klasifikasi. Pada penelitian ini dataset akan dibagi menjadi dua yaitu data latih dan data uji. Data latih berguna sebagai pembentuk dari model klasifikasi sedangkan data uji sebagai pengujian untuk mengevaluasi model yang sudah dibentuk.

Data yang berjumlah 1000 tersebut sudah diberi label sentimen secara manual yaitu 644 data berlabel positif, 140 data berlabel negatif, dan 216 data berlabel netral. Data kemudian di *resample* sehingga jumlah data berlabel positif, netral, dan negatif menjadi seimbang. Jumlah data yang sudah di *resample* menjadi 1932 data.

3.8. Pembobotan Kata (TF-IDF)

Pembobotan kata atau TF-IDF merupakan langkah yang dilakukan sebelum masuk ke dalam proses klasifikasi. Pembobotan kata berguna untuk melihat dan mengetahui seberapa penting suatu kata dari dokumen yang ada. Pembobotan kata bersifat statistik numerik.

3.9. Split Data (Pembagian Data)

Kategori dari dataset diberi secara manual ke dalam tiga kategori atau label yaitu positif, negatif dan netral. Pembagian data dalam penelitian ini yaitu 80% dari keseluruhan data sebagai data latih dan 20% sebagai data uji.

3.10. Klasifikasi

Pada proses klasifikasi ini dilakukan proses data training dan prediksi sentiment menggunakan algoritma *Naïve Bayes* dan *Decision Tree*. Data tersebut dibagi menjadi 3 bagian yaitu positif, netral dan negatif.

3.11. Akurasi

Pada tahap ini, akurasi data dihitung berdasarkan model yang telah dibuat. Pada tahap ini, metode evaluasi Confusion Matrix digunakan untuk melakukan perhitungan akurasi dari metode Naïve Bayes dan Decision Tree. Alur kerja metode ini adalah membandingkan label data uji yang telah ditentukan sebelumnya dengan label prediksi dari data uji yang dikelompokkan. pengelompokan. Hasil dari

penerapan metode ini adalah menemukan nilai error yang terjadi selama proses klasifikasi (Yulian, 2018).

3.12. Evaluasi

Evaluasi merupakan proses akhir dari semua proses pada penelitian ini. hasil evaluasi yang dilakukan berupa *accuracy*, *precision* dan *recall*. Sehingga dari hasil tersebut dapat dilihat perbandingan hasil klasifikasi dari algoritma *Naïve Bayes* dan *Decision Tree*.