

BAB III

METODE PENELITIAN

3.1 Subjek dan Objek Penelitian

3.1.1 Subjek Penelitian

Subjek penelitian yang digunakan adalah website Tokopedia untuk produk *skincare* skintific.

3.1.2 Objek Penelitian

Objek penelitian yang digunakan adalah data ulasan yang digunakan untuk analisis sentimen pada produk *skincare* skintific.

3.2 Alat dan Bahan

Alat dan Bahan yang digunakan dalam penelitian ini diantaranya yaitu :

3.2.1 Alat

Alat yang digunakan dalam penelitian ini terdiri dari dua jenis yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*), adapun perinciannya sebagai berikut :

1. Perangkat Keras (*Hardware*)

- a. Device : LAPTOP-DTMHT7M3
- b. Processor : AMD Ryzen 5 5500U with Radeon Graphics
2.10 GHz
- c. RAM : 8GB

2. Perangkat Lunak (*Software*)

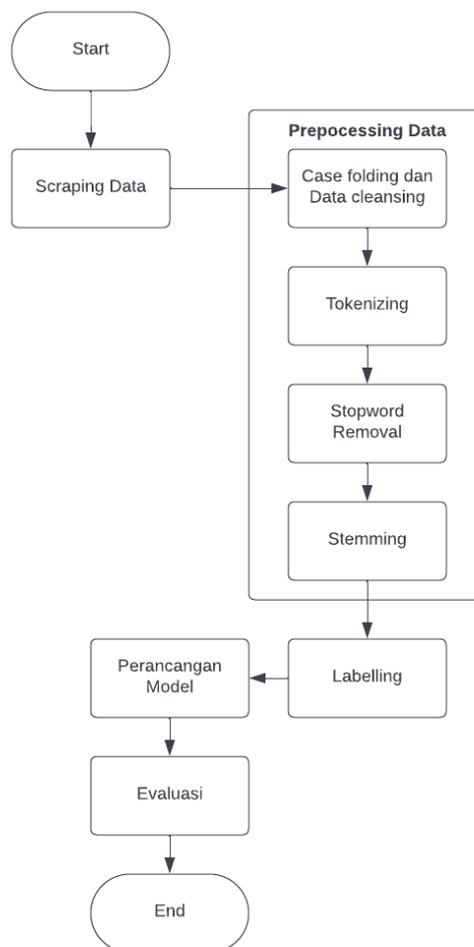
- a. Sistem Operasi : *Windows 11 Home Single Language*
- b. Bahasa Pemrograman : *Python*
- c. Aplikasi : *Google Colab*

3.2.2 Bahan

Bahan yang digunakan dalam penelitian ini adalah data *review* produk *skincare* yang diperoleh dari website tokopedia mengenai pengaruh *brand image* dan label harga pada produk *skincare* terhadap minat beli konsumen.

3.3 Diagram Alir Penelitian

Rancangan penelitian dalam Penyusunan Tugas akhir ini dapat dilihat pada diagram alir penelitian yang disajikan pada gambar 3.1.



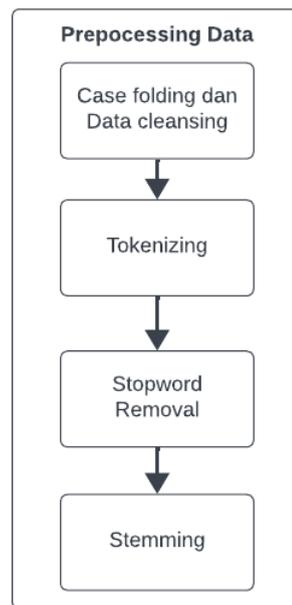
Gambar 3. 1 Diagram Alir Penelitian

3.3.1 *Scraping Data*

Scraping data diperoleh dari *review* produk skintific pada website tokopedia. Hasil dari *scraping* tersebut nantinya disimpan ke dalam file csv.

3.3.2 *Preprocessing Data*

Dataset tersebut kemudian dilakukan proses *preprocessing* untuk menangani noise yang ada pada data sehingga data yang dihasilkan menjadi terstruktur. Berikut *flowchart* pada *preprocessing* data pada gambar 3.2.



Gambar 3. 2 Flowchart Preprocessing Data

3.3.2.1 *Case Folding*

Pada tahap ini, dilakukan proses mengubah kalimat menjadi huruf kecil pada semua data.

3.3.2.2 *Data Cleansing*

Pada proses ini, dilakukan pembersihan data yang dapat mengganggu hasil analisis seperti karakter berulang, URL,

nama pengguna, tagar, simbol, angka, dan spasi tambahan, serta tanda baca.

3.3.2.3 *Stemming*

Stemming merupakan proses menghilangkan *afiks* seperti prefiks, konfiks, dan sufiks untuk mengubah kata berimbuhan menjadi ke bentuk akarnya.

3.3.2.4 *Tokenizing*

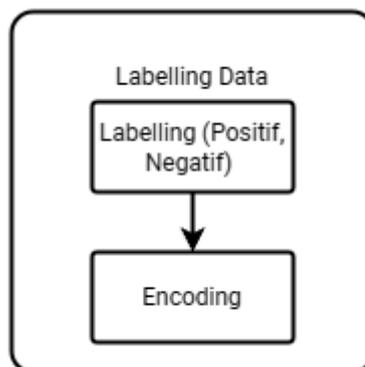
Tokenisasi merupakan proses pemecahan kalimat menjadi kata-kata sesuai dengan aturan bahasa.

3.3.2.5 *Filtering (Stopword Removal)*

Prosedur yang dikenal sebagai *Stopwords Removal* digunakan untuk menghilangkan katakata yang tidak berarti apa-apa.

3.3.3 *Labelling*

Tahap selanjutnya yaitu proses *labelling*. *Labelling* dilakukan untuk memperoleh nilai sentimen dari *review* produk oleh konsumen terhadap produk *skincare* skintific. Prose labelling dapat dilihat pada gambar 3.3.



Gambar 3. 3 Flowcart Labelling Data

3.3.4 Perancangan Model

Model yang digunakan pada penelitian ini adalah model Algoritma *Convolutional Neural Network*.

3.3.5 Pelatihan dan Pengujian Model

Pengujian model dilakukan dengan matriks konvolusi, seperti *precision*, *recall*, *f1-score* dan akurasi. Pada *confusion matrix* terdapat 4 hasil pengukuran, yaitu :

1. *True Negative (TN)*: Merupakan jumlah data *review* produk *skincare* skintific yang diklasifikasikan negatif dan terdeteksi dengan benar.
 2. *False Positive (FP)*: Merupakan jumlah data *review* produk *skincare* skintific yang diklasifikasikan negatif tetapi terdeteksi sebagai data klasifikasi positif.
 3. *True Positive (TP)*: Merupakan jumlah data *review* produk *skincare* skintific yang diklasifikasikan salah terdeteksi benar.
 4. *False Negative (FN)*: Merupakan kebalikan dari *True Positive*.
- 1) Akurasi : Menunjukkan hasil analisis data *review* yang diklasifikasi oleh model dengan tepat.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

- 2) Recall : Menunjukkan banyaknya data *review* untuk label positif yang terklasifikasi secara benar diantara data yang diprediksi sebagai positif.

$$Recall = \frac{TP}{TP+FN} \quad (3.2)$$

- 3) *Precision* : Merupakan nilai rasio antara jumlah data *review* berlabel positif yang terklasifikasi secara benar diantara semua data dengan label positif.

$$Precision = \frac{TP}{TP+FN} \quad (3.3)$$

- 4) *F-1 Score* : Nilai ini merupakan rata-rata dari *precision* dan *recall*

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (3.4)$$