

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Objek dan Subjek Penelitian

Subjek dari penelitian ini yaitu pengguna dari produk The Originote. dimana subjek akan melakukan komentar *review* produk di dalam situs YouTube. Objek penelitian ini adalah komentar para pengguna produk The Originote pada situs yaitu YouTube.

#### 3.2 Alat dan Bahan Penelitian

Penelitian ini menggunakan alat-alat dan bahan-bahan penelitian sebagai upaya menunjang jalanya penelitian yaitu sebagai berikut:

##### 3.2.1. Alat Penelitian

Alat yang digunakan pada penelitian ini dikelompokkan menjadi dua jenis yaitu perangkat lunak (*software*) dan perangkat keras (*hardware*), rincian dari perangkat keras yang digunakan tertera pada tabel 3. 1 sedangkan perangkat lunak yang digunakan dijabarkan pada tabel 3. 2.

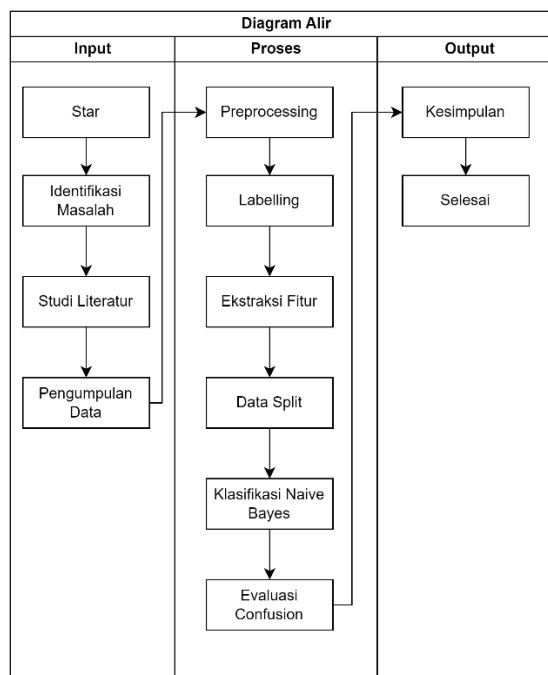
Tabel 3. 1 Tabel Perangkat Keras

No	Nama	Fungsi	Keterangan
1	Laptop	Digunakan untuk merancang dokumen proposal tugas akhir, merancang kode program, mencari referensi untuk menunjang progres penelitian.	Spesifikasi: 1. Intel Core i3-8665U 8.00 GB RAM. 2. <i>Operating System</i> : Windows 11

Tabel 3. 2 Tabel Perangkat Lunak

No	Nama	Fungsi	Keterangan
1	Google Collab	Perangkat lunak ini digunakan untuk menyusun kode program python.	Versi dari google collab yang digunakan
2	Web <i>Browser</i>	Perangkat lunak ini digunakan untuk mendukung jalanya penelitian seperti mencari referensi dan mengembangkan aplikasi.	<i>Browser</i> yang digunakan antara lain Chromium, Microsoft Edge, dan Mozilla Firefox

### 3.3 Diagram Alir Penelitian



Gambar 3. 1 Diagram Alir Penelitian

Gambar 3.1 diatas menjelaskan tentang diagram alir proses penelitian ini yang terdiri dari *input*, *process*, *output*. Berikut adalah penjelasan dari masing-masing tahapan penelitiannya:

### 3.3.1. Identifikasi Masalah

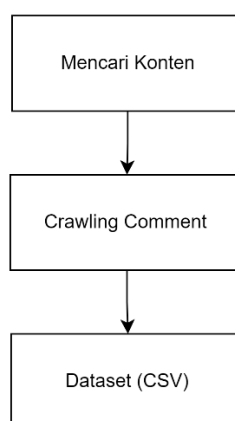
Tahapan identifikasi masalah adalah tahap penulis merumuskan masalah yang akan diangkat, tahap ini terdiri dari beberapa bagian. Perumusan latar belakang dibuat untuk dijadikan sebagai dasar penelitian yang dilakukan, metode yang akan digunakan juga sudah ditentukan. Selain latar belakang juga ditentukan rumusan masalah, pertanyaan, tujuan dan manfaat penelitian.

### 3.3.2. Studi Literature

Tahap studi literature adalah tahap penulis mengumpulkan data dari berbagai sumber seperti buku, jurnal, dan literatur sejenis yang berkaitan dengan topik penelitian. Data ini dijadikan sebagai referensi untuk mendukung penelitian, terutama dalam aspek seperti analisis sentimen, klasifikasi, teknik pengambilan data (scraping), dan metode Naïve bayes yang diterapkan dalam penelitian ini.

### 3.3.3. Pengumpulan Data

Tahap Teknik mengumpulkan data yang dilakukan adalah dengan mengumpulkan komen dari YouTube. Dalam penelitian ini tahapan sentimen analisis terdiri dari mencari konten pada YouTube, melakukan crawling data dengan API v3, mengubah menjadi .csv gambar 3.2.



Gambar 3. 2 Tahap pengumpulan data

### 3.3.4. *Preprocessing*

Salah satu hal yang harus dilakukan dalam pengolahan data teks adalah tahap pengolahan data menjadi lebih bersih atau dikenal dengan *preprocessing*. Tahap ini diperlukan karena data yang didapatkan tidak seluruhnya dapat digunakan dalam proses analisis, terdapat beberapa bagian yang dihilangkan atau dirubah. Dengan tahap ini maka data yang digunakan akan sesuai dengan kebutuhan pada proses analisis sehingga didapatkan hasil yang baik . Secara sistematis tahap *preprocessing* dilakukan dengan tahapan seperti berikut:

#### 1) *Case Folding*

*Case folding* adalah proses untuk mengkonversi semua karakter dalam dokumen menjadi huruf besar atau huruf kecil semua, agar mempercepat perbandingan selama proses pengindeksan.

#### 2) *Tokenizing*

*Tokenizing* merupakan proses pemisahan teks menjadi potongan kata yang disebut token. Tujuan dari proses ini adalah mendapatkan potongan kata yang akan menjadi entitas yang memiliki nilai dalam matriks dokumen teks yang akan dianalisis.

#### 3) *Normalization*

*Normalization* adalah proses teks yang dilakukan untuk memperbaiki kata kata yang salah eja atau kata yang disingkat. Misalnya, kata “tidak” dapat ditulis dengan “tdk, ngga, nggak, engga”, dan banyak lainnya. Dalam proses spelling normalization juga mengartikan bahasa asing khususnya english menjadi bahasa Indonesia.

#### 4) *Stopword Removal*

Proses ini dilakukan untuk menghilangkan kata-kata yang tidak penting karena seringnya muncul disetiap kometar review sehingga dapat mengganggu dalam proses pengenalan sentiment. Proses ini dilengkapi dengan menghilangkan hyperlink, case folding untuk mengubah seluruh kata menjadi seragam dalam huruf kecil, dan menghilangkan tanda bac a dengan regular expression.

### 5) *Stemming*

Proses mengubah kata yang memiliki imbuhan menjadi sebuah kata dasar. Dalam proses pengklasifikasian teks, banyak kata yang ditemukan memiliki imbuhan awalan atau akhiran dari sebuah kata. Penghapusan imbuhan pada penelitian ini dibantu dengan library sastrawi.

#### 3.3.5. *Pelabelan Data*

Proses analisis sentimen penelitian ini menggunakan metode Naive bayes. Metode Naive bayes membutuhkan label pada masing-masing baris datanya untuk pembangunan model. Pelabelan adalah tahap menentukan kalimat termasuk ke dalam kategori positif, negatif atau neutral. Tahap ini akan menggunakan metode *lexicon based* dengan menggunakan kamus lexicon berbahasa Inggris dari *library python TextBlob*. Konversi data dari bahasa Indonesia ke bahasa Inggris harus dilakukan karena perbedaan bahasa antara kamus lexicon yang digunakan dengan data yang akan dicari polaritasnya. *Library TextBlob* memiliki 2919 kata berbahasa Inggris dimana masing-masing kata sudah memiliki polaritas masing-masing. Dalam menentukan polaritas *TextBlob* akan melihat apakah kalimat tersebut memiliki skor positif, negatif atau netral [9].

### 3.3.6. Ekstra Fitur

Tahap pembobotan adalah tahap yang termasuk ke dalam tahap ekstraksi fitur. Tahap ini akan mengubah data teks menjadi data numerik yang dapat diolah oleh komputer. Pembobotan dalam penelitian ini dilakukan dengan metode Term Frequency Inverse Document Frequency (TF-IDF). TF-IDF akan memberikan nilai pada masing-masing kata pada setiap record data. Langkah-langkah dalam menentukan bobot tiap kata yaitu:

- 1) Menghitung frekuensi kemunculan kata pada setiap record data atau disebut sebagai *tf*.
- 2) Menghitung nilai IDF dengan persamaan:

$$idf = \log\left(\frac{n}{df}\right) \quad (3.1)$$

Keterangan:

$n$  = jumlah record data  $df$  = jumlah  
kemunculan kata pada  $n$

- 3) Menghitung nilai *tf-idf* pada masing-masing record data menggunakan persamaan:

$$tfidf = tf * idf \quad (3.2)$$

Keterangan:  $tf$  = kemunculan kata pada setiap record data

$idf$  = nilai *invers* dari kemunculan data

- 4) Lakukan langkah tiga untuk setiap kata pada masing-masing baris data.



