

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan pada platform media sosial Instagram @mie.gacoan, dan waktu penelitian ini berlangsung selama 3 bulan.

#### **3.2. Metode Pengumpulan Data**

Penelitian ini menggunakan metode pengumpulan data tambahan untuk melengkapi data yang telah ada. Berikut adalah tahapan metode yang dilakukan:

##### **3.2.1. Observasi**

Metode ini digunakan dengan cara mendapatkan data-data dan fakta dari pengamatan dengan cara *crawling* (mengambil data tertentu dari sebuah situs web) di Instagram @mie.gacoan untuk mendapatkan informasi dan pengetahuan yang diperlukan untuk bahan penelitian. Data yang digunakan dalam penelitian ini berupa sentimen pelanggan Mie Gacoan mengenai pelayanan yang telah diberikan oleh Mie Gacoan kepada pelanggan yang diambil dari bulan November 2023 - Januari 2024. Jumlah data komentar Instagram yang diambil ialah sebanyak 1156 data.

##### **3.2.2. Studi Pustaka**

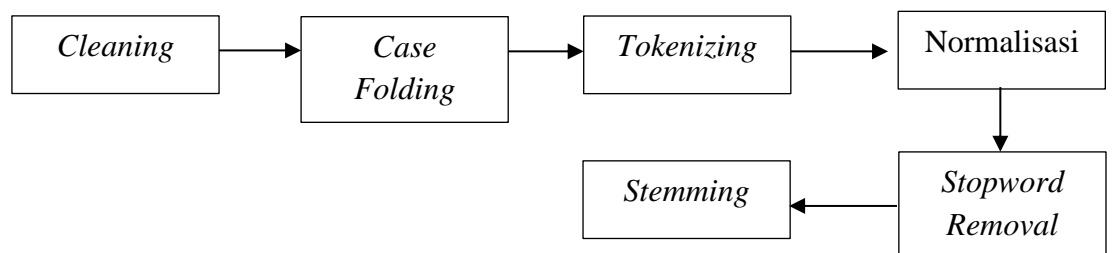
Dilakukan dengan cara mencari referensi berupa jurnal, buku, dan penelitian lain yang berkaitan dengan penelitian ini yang memiliki hubungan langsung dengan objek penelitian yang telah dipilih. Tujuan dari studi pustaka yaitu agar penulis memiliki landasan teori yang kuat.

### 3.3. Perangkat Penelitian

- a. Perangkat Lunak (*Software*)
  1. Microsoft Windows 10 Pro 64-bit adalah perangkat lunak sistem operasi yang digunakan
  2. Microsoft Office Excel 2016, software ini digunakan sebagai media penulisan datasheet.
  3. Jupyter Notebook, software ini digunakan sebagai media pengolahan dataset.
  4. *IGCommentsExport*, software ini digunakan untuk pengambilan dataset secara otomatis dari halaman web.
  
- b. Perangkat Keras (*Hardware*)
  1. Laptop Toshiba R634/M
  2. Intel Core i5-4319U 2.00GHz dan 2.6GHz
  3. 4 GB RAM
  4. Mouse

### 3.4. Praproses Data

Praproses data adalah langkah dimana teks diseragamkan dalam bentuk dan formatnya sehingga dapat dipersiapkan untuk diproses pada langkah berikutnya. Praproses teks mencakup beberapa langkah seperti, *Cleaning*, *Case Folding*, *Tokenizing*, *Normalisasi*, *Stopword Removal*, dan *Stemming*. Gambaran langkah-langkah *text preprocessing* dapat dilihat pada gambar 3.1.



**Gambar 3.1.** Tahapan *Text Preprocessing*

Keterangan:

1. *Cleaning*

*Cleaning* adalah tahapan menghilangkan atribut yang ada pada data seperti simbol, tanda baca, *hashtag*, URL, angka dan *noise* lainnya. Sehingga data menjadi lebih efisien dan dapat diolah dengan baik.

2. *Case Folding*

*Case Folding* adalah tahapan dimana semua huruf kapital pada teks komentar akan diubah ke dalam huruf kecil. Hal ini bertujuan agar dokumen pada teks komentar memiliki bentuk yang standar.

3. *Tokenizing*

*Tokenizing* adalah tahapan dimana kalimat akan dipecah menjadi unit-unit terkecil atau token. Tahap ini bertujuan untuk memudahkan pemrosesan dan analisis data.

4. Normalisasi

Normalisasi adalah tahapan penormalan bentuk kata yang kurang normal seperti singkatan, bahasa asing, dan kata yang kurang baku menjadi baku.

5. *Stopword Removal*

*Stopword Removal* adalah tahapan menghapus atau menghilangkan kata-kata yang tidak bermakna dalam dataset dan tidak memiliki pengaruh terhadap analisis sentimen.

6. *Stemming*

*Stemming* adalah tahapan mengembalikan suatu kata ke dalam bentuk kata dasar atau menghilangkan imbuhan (awalan dan akhiran) pada kata tersebut.

### 3.5. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah metode *Naïve Bayes Classifier*. Prosesnya dimulai dengan *crawling* Instagram @mie.gacoan menggunakan *tools IGCommentsExport*. Selanjutnya data yang sudah terkumpul akan diolah dengan menggunakan software *Python*.

Pendekatan yang diterapkan adalah dengan melakukan proses pelabelan secara manual pada data latih yang telah diolah terlebih dahulu dengan memasukkan data ke dalam kategori kelas positif dan negatif berdasarkan Kamus Besar Bahasa Indonesia (KBBI). Setelah proses pelabelan manual selesai, langkah selanjutnya adalah membuat model *Naive Bayes Classifier* dan menguji model tersebut menggunakan data uji. Hal ini dilakukan untuk mengetahui seberapa baik model dapat melakukan klasifikasi pada data.

### 3.6.Rancangan Pengujian Model

Metode yang digunakan dalam proses evaluasi adalah *confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya, *confusion matrix* menyajikan informasi perbandingan antara hasil klasifikasi yang diperoleh oleh sistem dengan hasil klasifikasi yang seharusnya terjadi. Seperti pada tabel 3.1 yang menggambarkan contoh *confusion matrix* prediksi dua kelas.

**Tabel 3.1.***Confusion Matrix*

| Fakta   | Prediksi              |                       |
|---------|-----------------------|-----------------------|
|         | Negatif               | Positif               |
| Negatif | <i>True Negative</i>  | <i>False Positive</i> |
| Positif | <i>False Negative</i> | <i>True Positive</i>  |

Keterangan:

1. *True Negative* (TN), yaitu jumlah data yang bernilai negatif dan diprediksi benar sebagai negatif.
2. *True Positive* (TP), yaitu jumlah data yang bernilai positif dan diprediksi benar sebagai positif.
3. *False Negative* (FN), yaitu jumlah data yang bernilai positif akan tetapi diprediksi sebagai data negatif.
4. *False Positive* (FP), yaitu jumlah data yang bernilai negatif akan tetapi diprediksi sebagai data positif.

Untuk mengukur performa klasifikasi *Naïve Bayes*, metode yang diterapkan melibatkan perhitungan akurasi (*accuracy*), presisi (*precision*), *recall*, dan *f1-score* dengan persamaan sebagai berikut ini:

- a. Akurasi (*accuracy*), yaitu jumlah prediksi yang benar pada dataset yang diperoleh dari tabel *confusion matrix*. Sehingga ditentukan pada persamaan berikut:

$$\text{Akurasi (accuracy)} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \times 100\%$$

- b. Presisi (*precision*), yaitu rasio prediksi antara jumlah positif yang benar dan salah. Sehingga dapat dihitung menggunakan persamaan berikut:

$$\text{Presisi (precision)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%$$

- c. *Recall*, yaitu rasio prediksi dari jumlah positif benar dan jumlah positif yang benar ditambah dengan jumlah negatif yang salah. Sehingga dapat dihitung melalui persamaan berikut:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

- d. *F1 Score*, yaitu usaha untuk menyeimbangkan nilai presisi dan *recall*. Sehingga dapat dihitung melalui persamaan berikut:

$$\text{F1 Score} = 2 \times \frac{(\text{Recall} \times \text{Presisi})}{(\text{Recall} + \text{Presisi})}$$