

## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Objek yang digunakan pada penelitian ini adalah sentimen dari masyarakat terhadap pinjol, adapun lokasi penelitian tersebut dilakukan melalui media sosial twitter dan implementasinya dilakukan di Universitas Buana Perjuangan Karawang. Penelitian akan dilaksanakan selama 6 bulan. Berikut rincian penelitian ditunjukkan pada tabel 3.1 dibawah ini.

Tabel 3.1 Jadwal Penelitian

| No | Kegiatan         | Bulan |   |   |   |   |   |
|----|------------------|-------|---|---|---|---|---|
|    |                  | 1     | 2 | 3 | 4 | 5 | 6 |
| 1  | Pengumpulan Data |       |   |   |   |   |   |
| 2  | Analisa Data     |       |   |   |   |   |   |
| 3  | Implementasi     |       |   |   |   |   |   |
| 4  | Penyajian        |       |   |   |   |   |   |

### 3.2 Prosedur Penelitian

Untuk prosedur penelitian dapat digambarkan dengan bentuk *flowchart* seperti dalam gambar dibawah ini :



Gambar 3.1 Prosedur Penelitian

Berdasarkan gambar Gambar 3.1 penelitian ini dimulai dengan pengumpulan data seperti studi literatur dan *crawling* data, lalu dilakukan analisis data hasil dari *crawling* data seperti pelabelan data dan *text processing*, kemudian hasil analisis data diimplementasi dengan *Term Frequency - Inverse Document Frequency* atau (*TF – IDF*) untuk menentukan bobot nilai pada kata dan dilakukan klasifikasi menggunakan algoritma *Naive Bayes* dengan mengelompokkan sentimen

positif, negatif dan netral. Data yang sudah diimplementasikan kemudian dilakukan pengujian dengan menggunakan *confusion matrix*.

### 3.3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dibagi menjadi dua, antara lain sebagai berikut:

#### 1. Studi Literatur

Studi Literatur yaitu menemukan landasan teori dan memperoleh data dari berbagai jurnal, buku dan internet, seperti definisi, contoh penelitian yang relevan, metode, data mining, analisis sentimen, dan data yang terkait dengan penelitian ini.

#### 2. *Crawling* Data

Data yang dikumpulkan diambil dari sebuah *tweet* pada media sosial twitter. Proses pengumpulan data dilakukan dengan menggunakan aplikasi RapidMiner dengan cara *crawling* data dengan sebuah *query* “pinjol”.

### 3.3.2 Analisis Data

Data hasil *crawling* merupakan data yang masih mentah dengan masih banyak kalimat-kalimat yang *noise* atau tidak terstruktur, untuk itu perlu dilakukan analisis data, adapun analisis data mempunyai tahapan sebagai berikut :

#### A. *Text Processing*

Sebelum memasuki tahap klasifikasi, diperlukan sebuah teks dokumen untuk menyiapkan data terlebih dahulu, proses tersebut dikenal dengan istilah *text processing*. *Text Processing* sendiri mempunyai manfaat untuk digunakan agar teks dalam sebuah data yang masih memiliki *noise* menjadi lebih terstruktur.

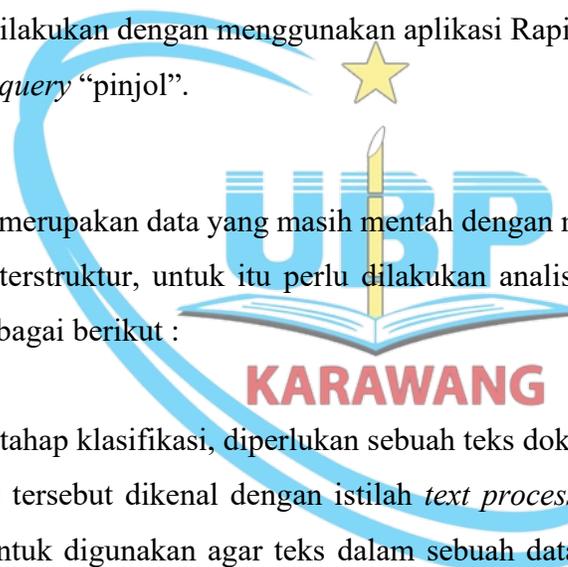
Ada beberapa tahapan dari *Text Processing* untuk digunakan pada penelitian ini, antara lain seperti : *Case Folding*, *Cleaning*, *Stopwords Removal*, *Stemming*, serta Normalisasi.

##### 1) *Case Folding*

*Case Folding* adalah sebuah proses untuk mengubah sebuah huruf kapital menjadi *lowercase* atau kecil semua. Proses tersebut dilakukan agar mempermudah pencarian, karena tidak semua dari dokumen teks konsisten dengan huruf kapital.

##### 2) *Cleaning*

*Cleaning* adalah sebuah proses untuk membersihkan data yang tidak diperlukan dari hasil *crawling*.



### 3) *Stopwords Removal*

*Stopwords Removal* adalah membuang kata-kata yang kurang penting atau menyimpan kosakata (menyimpan kata-kata penting). *Stopwords* adalah kata-kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak berarti. Contoh *stopword* bahasa Indonesia antara lain "yang", "dan", "di", "dari", dll. Maksud penggunaan *stopwords* adalah dengan menghilangkan kata-kata informasi rendah dari teks, kita dapat fokus pada kata-kata penting.

### 4) *Stemming*

*Stemming* adalah menghilangkan imbuhan pada kata, dan membuat kata yang ada imbuhan menjadi kata dasar.

### 5) Normalisasi

Normalisasi bertujuan mempermudah proses analisis sentimen terhadap entitas-entitas dikarenakan banyaknya istilah yg baku didalam sebuah *tweet* contohnya singkatan kata.

## B. Pelabelan Data

Data yang digunakan pada penelitian ini sebanyak 650 data, kemudian data tersebut diberikan suatu label secara manual untuk membentuk suatu analisis sentimen berupa positif, negatif dan netral.

### 3.3.3 Implementasi

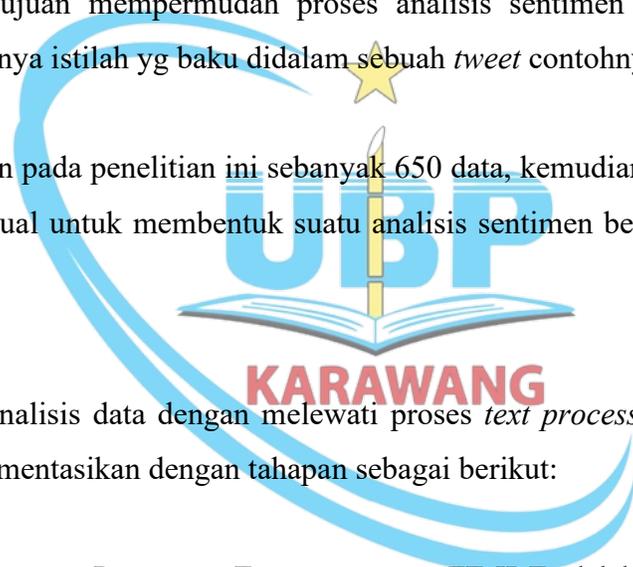
Setelah dilakukan analisis data dengan melewati proses *text processing* dan pelabelan data, kemudian data diimplementasikan dengan tahapan sebagai berikut:

#### 1. *TF – IDF*

*Term Frequency - Inverse Document Frequency* atau *TF-IDF* adalah metode algoritma yang umum digunakan untuk menghitung bobot setiap kata. Metode ini juga dikenal efisien, sederhana dan akurat. Metode ini akan menghitung nilai *term frequency (TF)* dan *inverse document frequency (IDF)* masing-masing tag (*word*) pada setiap dokumen di corpus. Secara sederhana, metode *TF-IDF* digunakan untuk mengetahui seberapa sering sebuah kata muncul dalam sebuah dokumen.

#### 2. Klasifikasi *Naive Bayes*

Setelah dilakukan pembobotan kata atau proses *TF-IDF*, kemudian data dilakukan klasifikasi dengan menggunakan algoritma *Naive Bayes*. Klasifikasi adalah metode untuk menemukan itemset dan mengelompokkannya sesuai dengan nilai bobot dari setiap item set. Data yang dilakukan pada proses klasifikasi yaitu 700 data dengan 80% sebagai data latih dan 20% sebagai



data uji. Dalam penelitian ini, data yang dikelompokkan dibagi menjadi tiga kategori, yaitu positif, netral dan negatif.

### 3.3.4 Pengujian

Pengujian yang dilakukan pada penelitian ini untuk mengevaluasi dan mengetahui performa dari algoritma (yang digunakan). Proses evaluasi dilakukan dengan menggunakan metode *confusion matrix* untuk menilai tingkat akurasi dari algoritma tersebut. Metode tersebut sangat membantu untuk menilai kinerja dari analisis klasifikasi (Ruger *et al.*, 2021).

Pengujian tersebut dilakukan dengan cara menghitung tingkat *accuracy*, *recall*, dan *precision* yang dapat ditampilkan dalam bentuk persentase.

1. *Accuracy*

*Accuracy* atau akurasi adalah hasil dari skala prediksi yang akurat.

2. *Precision*

*Precision* merupakan suatu jumlah teks yang signifikan dalam kalangan dokumen yang sebelumnya disortir menggunakan sebuah sistem.

3. *Recall*

*Recall* merupakan suatu jumlah teks relevan dalam kalangan dokumen teks dari koleksi.

Adapun yang digunakan untuk mengukur *accuracy*, *precision*, dan *recall* seperti tabel 3.2 dan persamaan berikut ini:

Tabel 3. 2 *Confusion Matrix*

| Manual  | Sistem              |                    |                     |
|---------|---------------------|--------------------|---------------------|
|         | Positif             | Netral             | Negatif             |
| Positif | True Positive (TP)  | False Neutral (FN) | False Negative (FN) |
| Netral  | False Positive (FP) | True Neutral (TN)  | False Negative (FN) |
| Negatif | False Positive (FP) | False Neutral (FN) | True Negative (TN)  |

Sumber : Eza Ananda Putra (2021)

$$Accuracy = \frac{TP+TN+TN}{TP+TN+TN+FP+FN+FN} * 100\% \quad (2)$$

$$Precision = \frac{TP}{FP+TP} * 100\% \quad (3)$$

$$Recall = \frac{TP}{FN+FN+TP} * 100\% \quad (4)$$

Dimana :

- TP adalah *True* Positif, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TNt adalah *True* Netral, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- TN adalah *True* Negatif, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FP adalah *False* Positif, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.
- FNt adalah *False* Netral, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah *False* Negatif, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

### 3.3 Peralatan Penelitian

Pada peralatan penelitian terdiri dari perangkat keras dan perangkat lunak.

#### 3.3.1 Perangkat Keras

Perangkat keras yang digunakan pada penelitian ini antara lain:

- Processor : Intel Core i5-9300H
- RAM : 8 GB DDR4
- HDD : 256 GB

#### 3.3.2 Perangkat Lunak

Perangkat lunak yang digunakan pada penelitian ini antara lain:

- Operating system Windows 10
- Microsoft Office 2010
- Mozilla Firefox
- RapidMiner Studio
- Anaconda

