# BAB III

# METODE PENELITIAN

## 3.1. Objek Penelitian

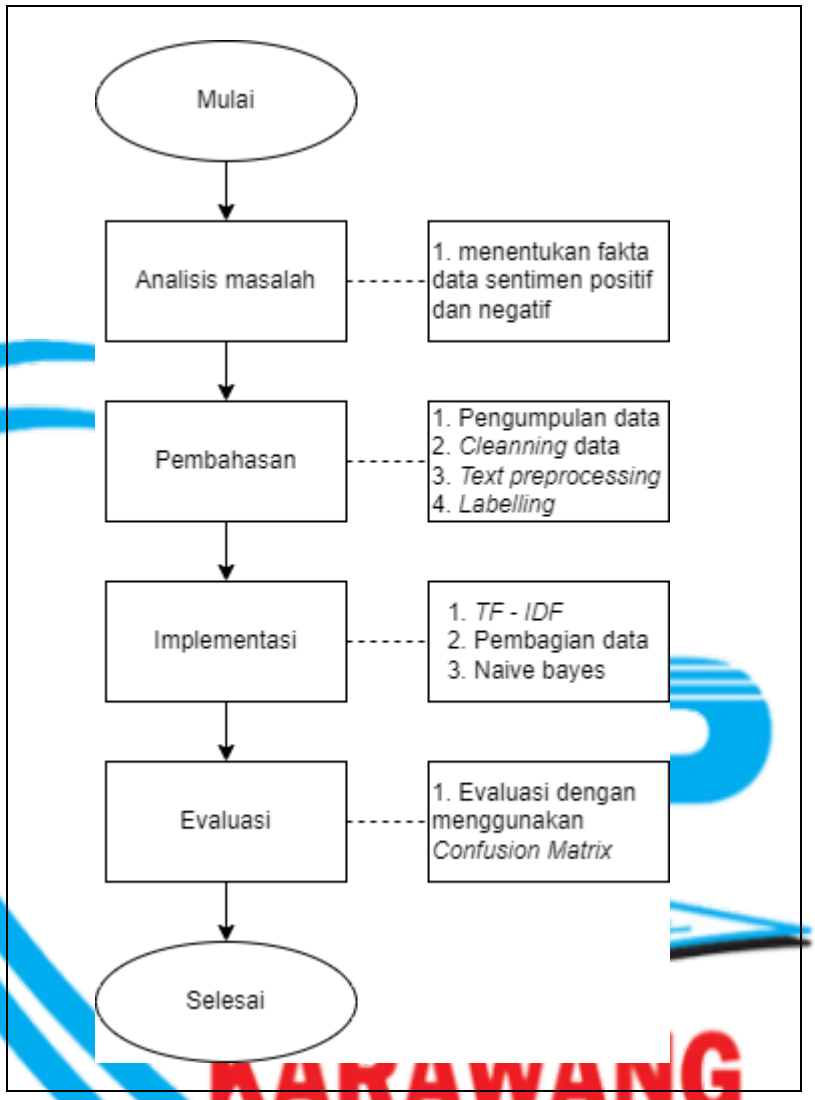
Objek penelitian ini adalah analisis sentimen larangan ekspor nikel. Adapun sumber data yang digunakan dari media sosial twitter. Penelitian ini dilakukan di Laboratorium Riset Universitas Buana Perjuangan Karawang selama 7 bulan.

Tabel 3. 1 Jadwal Penelitian

| No | Kegiatan | Bulan | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | Analisis masalah | | | | | | | |
| 2 | Pembahasan | | | | | | | |
| 3 | Implementasi | | | | | | | |
| 4 | Evaluasi | | | | | | | |

## 3.2. Prosedur Penelitian

Tahapan prosedur penelitian dimulai dari Analisis masalah, Pembahasan, Implementasi dan Evaluasi. Berikut Prosedur penelitian pada Gambar 3.1




Gambar 3. 1 Prosedur Penelitian

Berdasarkan Gambar 3.1 Penelitian dimulai dari Pembahasan yang terdiri dari pengumpulan data, *Cleanning* data, *Text preprocessing* dan *labelling* selanjutnya implementasi yang terdiri dari TF – IDF, Pembagian data dan Algoritma Naïve Bayes dan yang terakhir adalah evaluasi dengan menggunakan *confusion matrix*.

### 3.2.1. Analisis masalah

Proses pencarian permasalahan dilakukan dengan tujuan untuk mengetui fakta dari data yang sebenarnya. Cara menentukan fakta data analisis sentimen yang bernilai positif dan negatif diantanya:

1. Sentimen positif dapat diartikan sebagai suatu tindakan atau tanggapan masyarakat yang dilakukan dengan tidak adanya penolakan atau bersifat mendukung.
2. Sentimen negatif dapat diartikan sebagai suatu tindakan atau tanggapan yang dilakukan dengan adanya penolakan atau bersipat tidak mendukung.

### 3.2.2. Pembahasan

Pembahasan Memiliki Berbagai tahapan yaitu Pengumpulan Data, *Remove* Data, *Text Preprocessing* dan *Labelling* diantaranya sebagai berikut:

1. Pengumpulan Data

Tahap pertama merupakan pengumpulan data Tahapan ini dilakukan dengan mempelajari buku dan jurnal karya ilmiah yang berkaitan dengan *data mining*. Konsep yang dipelajari merupakan Analisis sentiment terhadap larangan ekspor nikel oleh pemerintah Indonesia dengan menggunakan algoritma Naïve Bayes Dataset yang akan digunakan pada analisis sentiment menggunakan algoritma Naïve Bayes. Data yang diolah merupakan data sentimen yang berasal dari media sosial twitter. Dengan Menggunakan program khusus yaitu twint.

```
$ twint -s "bunga" -o bunga.csv --limit 10 --since 2019-07-01 --until 2022-08-24 --csv
```

Gambar 3. 2 Contoh pengambilan data

Berdasarkan Gambar 3.2 Tahapan untuk pengambilan data yaitu dengan menggunakan liblary twint yaitu dengan memasukan pencarian bunga lalu akan menyimpan secara otomatis dengan nama bunga.csv dengan jumblah total data yaitu 10 dan data di ambil dari 01 Juli 2019 sampai 24 Agustus 2022.

2. *Cleanning* Data

Tahap selanjutnya adalah *Cleanning data* tahapan dari penghapusan Data yang berupa Data Duplikat dan Bahasa asing dari hasil data yang telah didapatkan pada tahapan pengambilan data, data akan dihapus secara manual.

3. *Text preprocessing*

Sebelum dilakukannya proses *klasifikasi*, text dokumen atau data harus disiapkan terlebih dahulu, proses tersebut biasanya dinamakan dengan *text processing*. Tahapan *text processing* berguna untuk text yang terdapat banyak noise atau tidak terstruktur menjadi terstruktur. *Text processing* memiliki beberapa tahapan yaitu *case folding, Konversi slangword, Stopword removal, Steamming* dan *Tokenizing.*

a. *Case Folding*

Proses *Case Folding* untuk merubah semua huruf didalam data teks dari huruf "a" sampai "z" menjadi huruf kecil dan menghapus berbagai elemen. Berikut Contoh *Case Folding* pada Tabel 3.2.

Tabel 3. 2 Contoh hasil *Case Folding*

| | |
|---|---|
| *Input* | BUNGA MAWAR SANGAT WANGI |
| Proses | data['KOMENTAR']=data['KOMENTAR'].str.lower()<br>data['casefold']=data['KOMENTAR'].apply(casef)<br>`def casef(text):`<br>`text = text.replace('\\t'," ").replace('\\n'," ").replace('\\u'," ").replace('\\',"")`<br>`text = text.encode('ascii', 'replace').decode('ascii')`<br>`text = ' '.join(re.sub("([@#][A-Za-z0 9]+)\|(\w+:\/\/\S+)"," ", text).split())`<br>`text= text.replace("http://", " ").replace("https://", " ")`<br>`text = re.sub(r"\d+", "", text)`<br>`text.translate(str.maketrans("","",string.punctuation))`<br>`text = text.strip()`<br>`text = re.sub('\s+',' ',text)`<br>`return text`<br>hasil_casefolding=data['casefold']<br>hasil_casefolding |
| *Output* | bunga mawar sangat wangi |

Berdasarkan Tabel 3.2 Tahapan untuk mengubah semua huruf besar menjadi huruf kecil dan menghapus berbagai elemen.

b. *Tokenizing*

*Tokenizing* berfungsi untuk memecah komentar menjadi satuan kata.Proses yang berurutan untuk memecah komentar atau kata. Berikut Contoh *Tokenizing* pada Tabel 3.3.

Tabel 3. 3 Contoh hasil *Tokenizing*

| | |
|---|---|
| *Input* | BUNGA MAWAR SANGAT WANGI |
| Proses | data['komen_tokens']=data['casefold']<br>.apply(word_tokenize_wrapper) |
| *Output* | [bunga, mawar, sangat, wangi] |

c. *Konversi Slangword*

*Konversi Slangword* digunakan untuk proses mengubah kata tidak baku ke dalam kata baku, Tahap ini dilakukan dengan menggunakan bantuan kamus slangword. Kata yang terdapat pada kamus slangword nantinya akan diubah menjadi kata baku. Berikut Contoh *Konversi Slangword* pada Tabel 3.4.

Tabel 3. 4 Contoh hasil *Konversi Slangword*

| | |
|---|---|
| *Input* | BUNGA MAWAR SANGAT WANGI |
| Proses | normalized_word=pd.read_excel("kamus perbaikan kata.xlsx")<br>data['komen_perbaikan']=data['komen_tokens']<br>.apply(normalized_term) |
| *Output* | [bunga, mawar, sangat, harum] |

Berdasarkan Tabel 3.4 Tahapan untuk merubah kata baku menjadi tidak baku yaitu dengan mengguanakan kamus perbaikan data.xlsx, kamus perbaikan data dibuat berdasarkan kata tidak baku dan baku.

d. *Stopwords removal*

*Stopwords removal* adalah proses penghapusan kata-kata umum atau kata-kata yang tidak relevan dalam teks. Kata-kata umum ini biasanya termasuk dalam kategori kata bantu (*stopwords*).

Tabel 3. 5 Contoh hasil *Stopwords removal*

| | |
|---|---|
| *Input* | BUNGA MAWAR SANGAT WANGI |
| Proses | normalized_word=pd.read_excel("kamus perbaikan kata.xlsx")<br>data['komen_perbaikan']=data['komen_tokens']<br>.apply(normalized_term) |
| *Output* | [bunga, mawar, sangat, harum] |

Berdasarkan Tabel 3.5 Tahapan *Stopwords removal* yaitu digunakan untuk menghapus kata yang tidak relevan.

e. *Steamming*

*Steamming* digunakan untuk proses pengubahan bentuk kata menjadi kata dasar atau tahap mencari root kata dari tiap kata hasil filtering. Berikut Contoh *Steamming* pada Tabel 3.6.

Tabel 3. 6 Contoh hasil *steamming*

| | |
|---|---|
| *Input* | BUNGA MAWAR SANGAT WANGI |
| Proses | data['komen_stemmed']=data['komen_filtered']<br>.apply(get_stemmed_term) |
| *Output* | [bunga, mawar, sangat, harum] |

Berdasarkan Tabel 3.6 Tahapan stemming yaitu digunakan untuk mencari komen hasil dari filtering lalu akan diambil berdasarkan kata dasarnya.

*4. Labelling*

Tahap selanjutnya adalah *Labelling* tahapan berisi hasil dari analisis sentimen yang berupa hasil positif dan negatif data yang telah didapatkan pada tahapan *Text Preprocessing*, Pelabelan akan dilakukan secara manual.

3.2.3. Implementasi

1. TF-IDF

Setelah melakukan tahapan *text processing* kemudian akan dilakukan pembobotan data menggunakan TF-IDF berikut adalah tahapan TF-IDF

1. Hitung *term frequency* (TF) Jumlah kemunculan istilah dalam dokumen dibagi dengan jumlah total kata dalam dokumen.

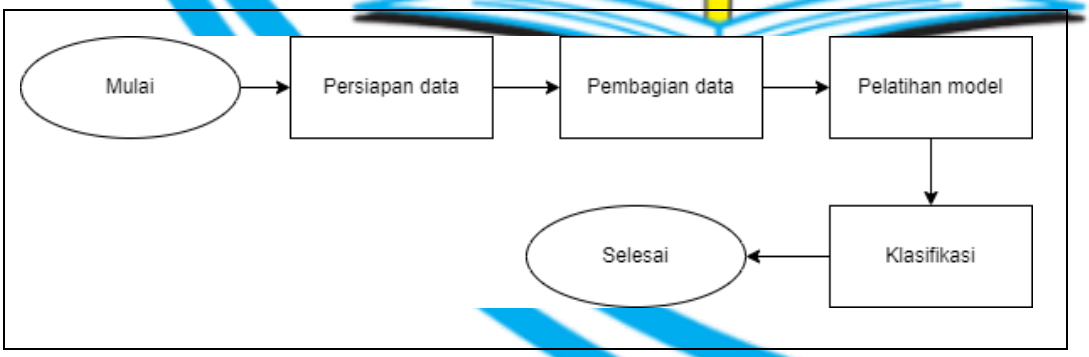
2. Hitung frekuensi dokumen terbalik (IDF) Logaritma dari jumlah total dokumen dalam korpus dibagi dengan jumlah dokumen yang berisi term.
3. Hitung TF-IDF Hasil perkalian TF dan IDF. Nilai TF-IDF menunjukkan pentingnya istilah dalam dokumen.
4. Normalisasi TF-IDF Pastikan nilai TF-IDF berada pada skala yang sama sehingga dapat dibandingkan di seluruh dokumen.

2. Pembagian Data

Setelah melakukan tahapan TF-IDF kemudian akan dilakukan Pembagian data, Pada proses ini data akan dibagi menjadi dua yaitu data latih sebanyak 80% dan data uji sebanyak 20% secara acak.

3. Naïve Bayes

Setelah melakukan tahapan TF-IDF, kemudian data akan diimplementasikan menggunakan algoritma Naïve bayes. Tahapannya dimulai dengan mempersiapkan data yang sudah di dapatkan media sosial twitter. Dalam penelitian ini data yang dikelompokan dibagi menjadi dua bagian yaitu positif dan negatif. Berikut tahapan algoritma Naïve Bayes pada Gambar 3.8.




Gambar 3. 3 Tahapan Algoritma Naïve Bayes

1. Persiapan data: Tahap pertama mempersiapkan data untuk digunakan.

2. Pembagian data: Setelah data siap, pisahkan menjadi dua bagian yaitu data latih dan data uji.

3. Model pelatihan: Tahap selanjutnya adalah melatih model Naive Bayes menggunakan data uji.

4. Klasifikasi: Setelah model dilatih, model tersebut dapat digunakan untuk mengklasifikasikan objek baru dan memberikan kelas tersebut probabilitas tertinggi.

3.2.4. Evaluasi

Evaluasi dilakukan dengan menggunakan *Confusion Matrix* untuk menghitung nilai accuracy, precision dan recall yang akan ditampilkan dalam bentuk presentase.

1. *Accuracy*

*Accuracy* merupakan presentase dari total sentimen yang benar dikenal, untuk menghitung nilai accuracy digunakan persamaan dibawah ini :

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

Keterangan :

TP : Jumlah positif yang diklasifikasi sebagai positif

TN : Jumlah negatif yang diklasifikasi sebagai negatif

FP : Jumlah negatif yang diklasifikasi sebagai positif

FN : Jumlah positif yang diklasifikasi sebagai negatif.

2. *Precision*

*Precision* merupakan perbandingan antara informasi penting yang ditemukan dengan jumlah informasi yang ditemukan. Persamaan berikut digunakan untuk menghitung nilai presisi:

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

Keterangan :

TP : Jumlah positif yang diklasifikasi sebagai positif

FP : Jumlah negatif yang diklasifikasi sebagai positif

3. *Recall*

*Recall* merupakan perbandingan jumlah materi relevan yang ditemukan. Untuk menghitung *recall* digunakan persamaan ini:

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

Keterangan :

TP : Jumlah positif yang diklasifikasi sebagai positif

FN : Jumlah positif yang diklasifikasi sebagai negatif.