

## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Objek penelitian yang digunakan berupa data *tweet* yang berkaitan dengan *cryptocurrency* pada media sosial twitter yang ditunjukkan pada gambar 3.1 dan 3.2.



Gambar 3.1 Postingan Positif



Gambar 3.2 Postingan Negatif

Gambar diatas merupakan sampel postingan sentimen positif dan negatif dari pengguna twitter mengenai *cryptocurrency*. Postingan dengan sentimen positif ditunjukkan oleh gambar 3.1 yang terdapat kalimat “crypto itu bukan judi, lebih investasi”. Sementara sentimen negatif ditunjukkan oleh gambar 3.2 pada kalimat “hasil kerugian dari kebodohan” dan “ketipu mau convert dollar, main kripto, dll”.

### 3.2 Lokasi Dan Waktu Penelitian

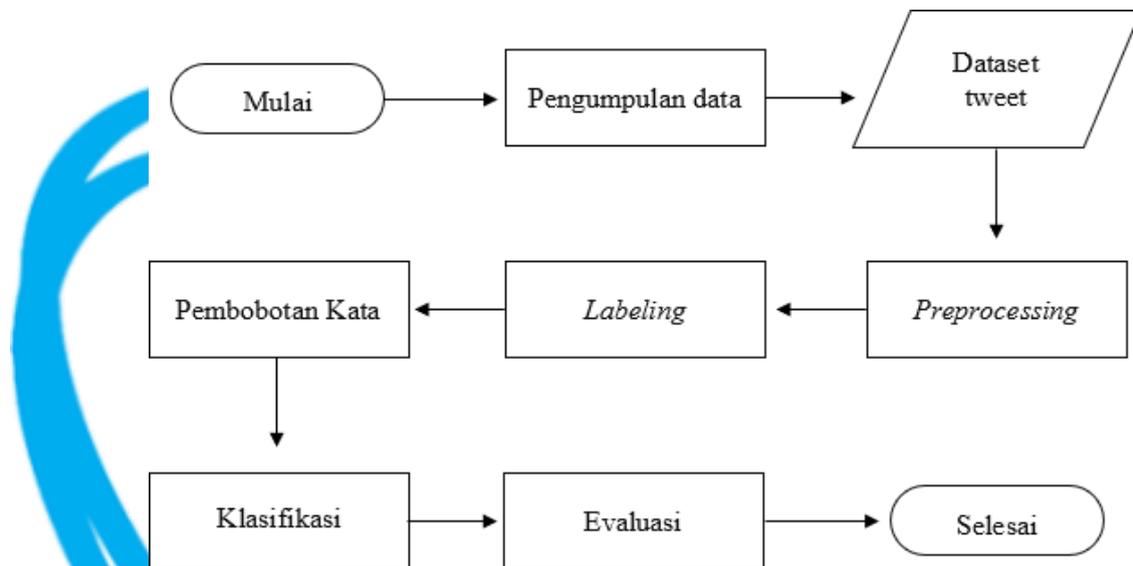
Lokasi kegiatan penelitian dilaksanakan di Laboratorium Riset Universitas Buana Perjuangan Karawang dimulai sejak bulan Januari 2022. Jadwal kegiatan penelitian ditunjukkan pada tabel 3.1.

Tabel 3.1 Jadwal Penelitian

Kegiatan	Februari	Maret	April	Mei	Juni	Juli	Agustus
Studi Literatur							
Penulisan Proposal							
Pengumpulan Data							
Preprocessing							
Labeling							
Pembobotan Kata							
Klasifikasi							
Pengujian dan Evaluasi							
Penulisan Laporan							

### 3.3 Prosedur Penelitian

Prosedur penelitian merupakan gambaran umum dari alur penelitian yang dilaksanakan selama pengerjaan tugas akhir. Prosedur penelitian dalam bentuk diagram alir ditunjukkan pada gambar 3.3.



Gambar 3.1 Prosedur Penelitian

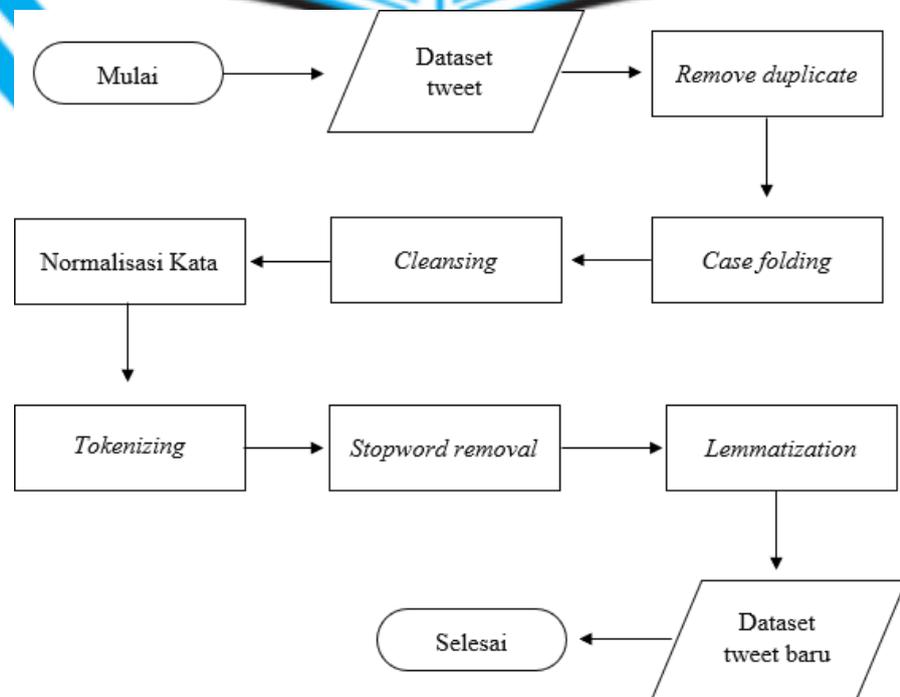
Berdasarkan gambar 3.3. Penelitian ini diawali dengan tahap pengumpulan data menggunakan teknik *scraping* dengan bahasa pemrograman python. Data yang diambil berupa kumpulan *tweet* berbahasa Indonesia dari media sosial Twitter yang berkaitan dengan *cryptocurrency*. Tahap selanjutnya setelah pengambilan data yaitu *preprocessing* yang berguna untuk menyeleksi dan membersihkan data mentah dari beberapa bagian yang dapat mengganggu proses selanjutnya. Setelah data menjadi bersih lalu dilakukan labeling ke dalam dua kelas yaitu positif dan negatif menggunakan python dengan *library* *textblob*. Setelah tahap *labeling* yaitu pembobotan kata menggunakan metode TF-IDF yang dibarengi dengan implementasi algoritma yang akan digunakan yaitu naive bayes. Tahap terakhir yaitu evaluasi yang bertujuan untuk mengetahui hasil dari pengujian yang sudah dilakukan, pada tahap ini menggunakan metode *confusion matrix* yang telah digunakan pada penelitian sebelumnya.

### 3.3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara mengumpulkan *tweet* berbahasa Indonesia pada media sosial Twitter yang membahas mengenai *cryptocurrency*. Pengumpulan data dilakukan pada media sosial twitter karena topik yang berkaitan dengan *cryptocurrency* sering kali masuk dalam trending Twitter Indonesia. Data dikumpulkan secara acak baik dari *tweet* pengguna biasa atau dari media *online*. Proses *scraping* dilakukan menggunakan bahasa pemrograman python.

### 3.3.2 Preprocessing

Data *tweet* yang sudah dikumpulkan dari hasil *scraping* tidak dapat langsung digunakan untuk klasifikasi karena masih terdapat banyak gangguan, karena itu memerlukan tahap bernama *preprocessing*. Tujuan dari tahap *preprocessing* yaitu membersihkan data atau *tweet* dari kata-kata yang tidak diperlukan dalam penelitian dan tidak memiliki makna serta mengurangi volume kata. Adapun tahapan dari *preprocessing* ditunjukkan pada gambar 3.4.



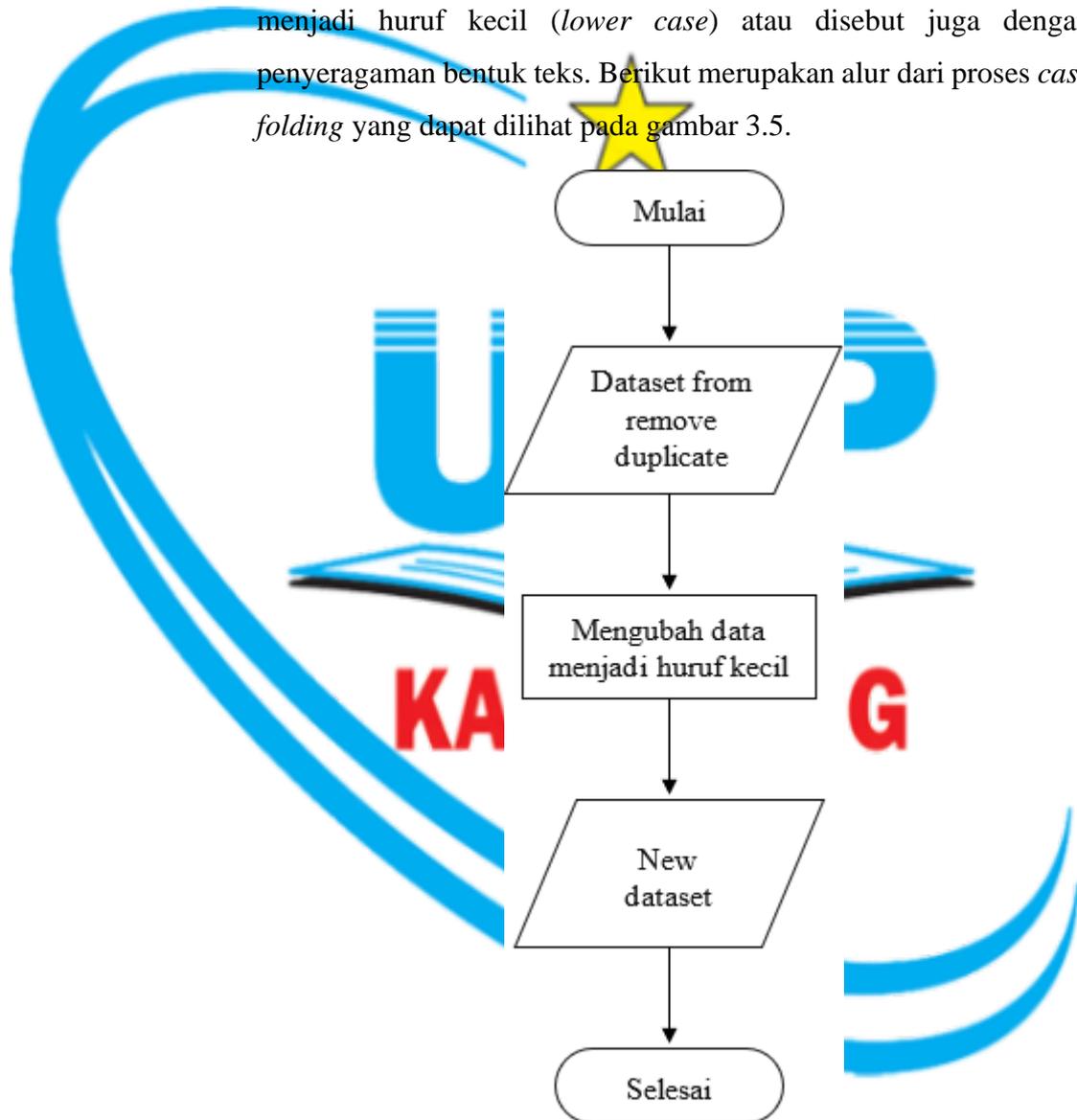
Gambar 3.2 Tahapan *Preprocessing*

a. *Remove Duplicate*

*Tweet* yang sama akan dihapus sehingga hanya *tweet* unik yang diproses pada langkah berikutnya, proses akan lebih cepat karena data sudah berkurang.

b. *Case Folding*

*Case folding* merupakan proses mengubah semua huruf dalam teks menjadi huruf kecil (*lower case*) atau disebut juga dengan penyeragaman bentuk teks. Berikut merupakan alur dari proses *case folding* yang dapat dilihat pada gambar 3.5.



Gambar 3.3 *Flowchart case folding*

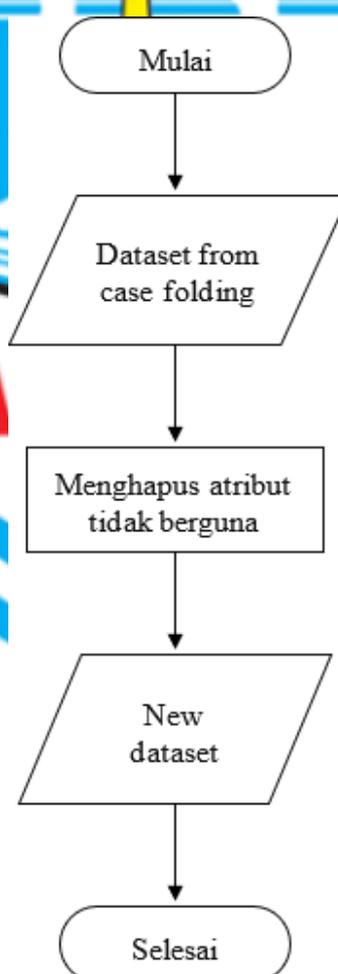
Sebagai gambaran proses *case folding*, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.2.

Tabel 3.2 Contoh proses *case folding*

Sebelum	Sesudah
Dan investasi itu bkn cm tanah, rumah, saham, reksadana aja. Tp bisa juga berupa aset kripto #TokocryptoAja#UpgradeYuk @Tokocrypto	dan investasi itu bkn cm tanah, rumah, saham, reksadana aja. tp bisa juga berupa aset kripto #tokocryptoaja#upgradeyuk @tokocrypto

c. *Cleansing*

*Cleansing* merupakan proses menghilangkan atribut yang dapat mengganggu proses klasifikasi seperti url, simbol, *username*, *mention* dan *emoticon*. Berikut merupakan alur dari proses *cleansing* yang dapat dilihat pada gambar 3.6.

Gambar 3.4 *Flowchart cleansing*

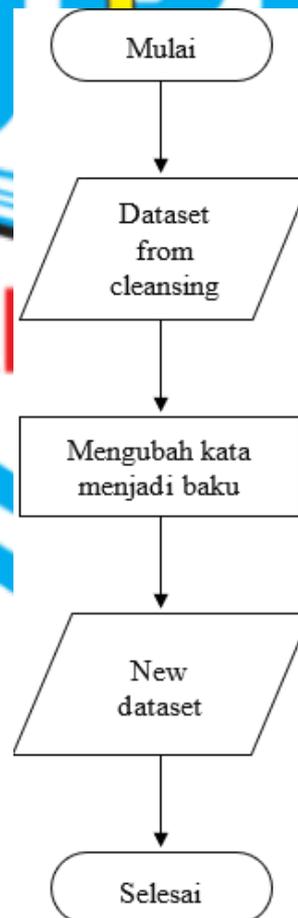
Sebagai gambaran proses *cleansing*, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.3.

Tabel 3.3 Contoh proses *cleansing*

Sebelum	Sesudah
dan investasi itu bkn cm tanah, rumah, saham, reksadana aja. tp bisa juga berupa aset kripto #tokocryptoaja#upgradeyuk @tokocrypto	dan investasi itu bkn cm tanah rumah saham reksadana aja tp bisa juga berupa aset kripto

d. Normalisasi Kata

Normalisasi kata merupakan proses memodifikasi kata-kata yang tidak sesuai dengan EYD. Berikut merupakan alur dari proses normalisasi kata yang dapat dilihat pada gambar 3.7.



Gambar 3.5 *Flowchart* Normalisasi Kata

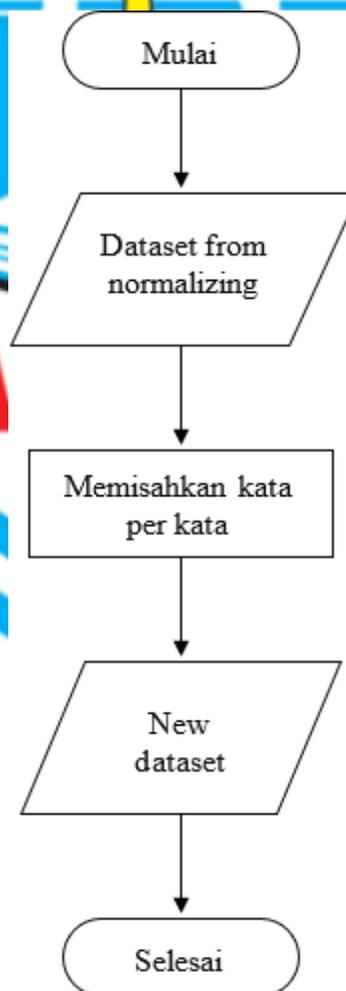
Sebagai gambaran proses normalisasi kata, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.4.

Tabel 3.4 Contoh proses normalisasi kata

Sebelum	Sesudah
dan investasi itu bkn cm tanah rumah saham reksadana aja tp bisa juga berupa aset kripto	dan investasi itu bukan cuman tanah rumah saham reksadana saja tapi bisa juga berupa aset kripto

e. *Tokenizing*

*Tokenizing* merupakan proses untuk memecah kalimat menjadi potongan kata tunggal. Berikut merupakan alur dari proses *tokenizing* yang dapat dilihat pada gambar 3.8.



Gambar 3.6 *Flowchart tokenizing*

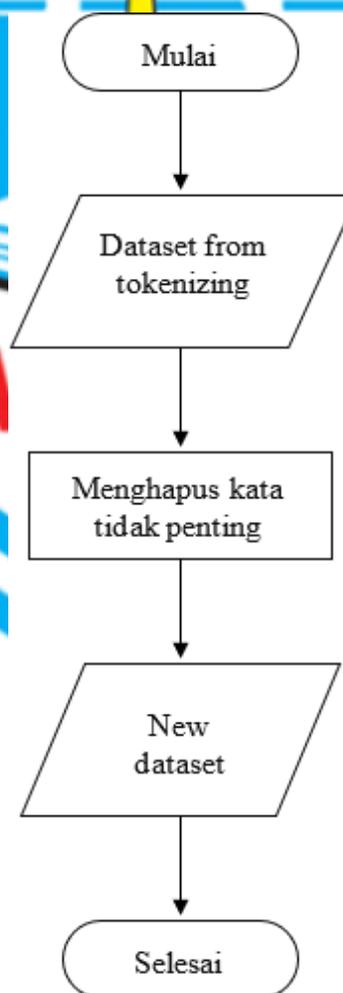
Sebagai gambaran proses *tokenizing*, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.5.

Tabel 3.5 Contoh proses *tokenizing*

Sebelum	Sesudah
dan investasi itu bukan cuman tanah rumah saham reksadana saja tapi bisa juga berupa aset kripto	[dan, investasi, itu, bukan, cuma, tanah, rumah, saham, reksadana, saja, tapi, bisa, juga, berupa, aset, kripto]

f. *Stopword Removal*

*Stopword Removal* merupakan proses menghilang kata-kata yang kurang penting dalam proses analisis. Berikut merupakan alur dari proses *stopword removal* yang dapat dilihat pada gambar 3.9.



Gambar 3.7 Flowchart *stopword removal*

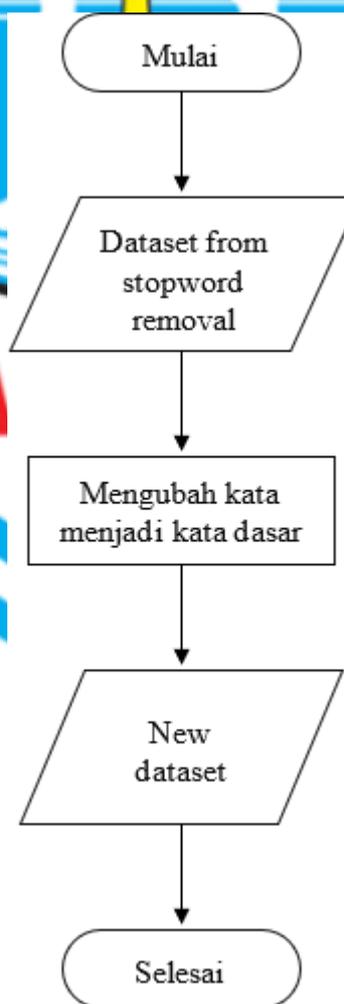
Sebagai gambaran proses *stopword removal*, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.6.

Tabel 3.6 Contoh proses *stopword removal*

Sebelum	Sesudah
[dan, investasi, itu, bukan, cuma, tanah, rumah, reksadana, saja, tapi, bisa, juga, berupa, aset, kripto]	[investasi, tanah, rumah, saham, reksadana, aset, kripto]

g. *Lemmatization*

*Lemmatization* merupakan proses mengubah suatu kata menjadi kata dasar. Berikut merupakan alur dari proses *lemmatization* yang dapat dilihat pada gambar 3.10.



Gambar 3.8 Flowchart *lemmatization*

Sebagai gambaran proses *lemmatization*, berikut merupakan contoh *tweet* yang dapat dilihat pada tabel 3.6.

Tabel 3.7 Contoh proses *lemmatization*

Sebelum	Sesudah
[manfaat, aset, kripto, iya, media, pertukaran]	['manfaat', 'aset', 'kripto', 'iya', 'media', 'tukar']

### 3.3.3 Labeling

Langkah selanjutnya setelah *preprocessing* yaitu melakukan *labeling*. Proses ini dilakukan untuk memberi label pada konten atau *tweet* pengguna twitter dan kemudian setiap *tweet* akan diberi indeks untuk membedakan setiap nilai sentimen. Dalam penelitian ini terdapat dua kelas dalam memberi label ke suatu *tweet* yaitu kelas positif dan negatif. Hasil dari pelabelan akan di visualisasikan ke dalam dua bentuk yaitu *bar chart* dan *word cloud*. *Bar chart* digunakan untuk menunjukkan perbandingan pada jumlah sentimen sedangkan *word cloud* akan menampilkan kata-kata umum yang sering muncul pada hasil sentimen positif dan negatif.

### 3.3.4 Pembobotan Kata

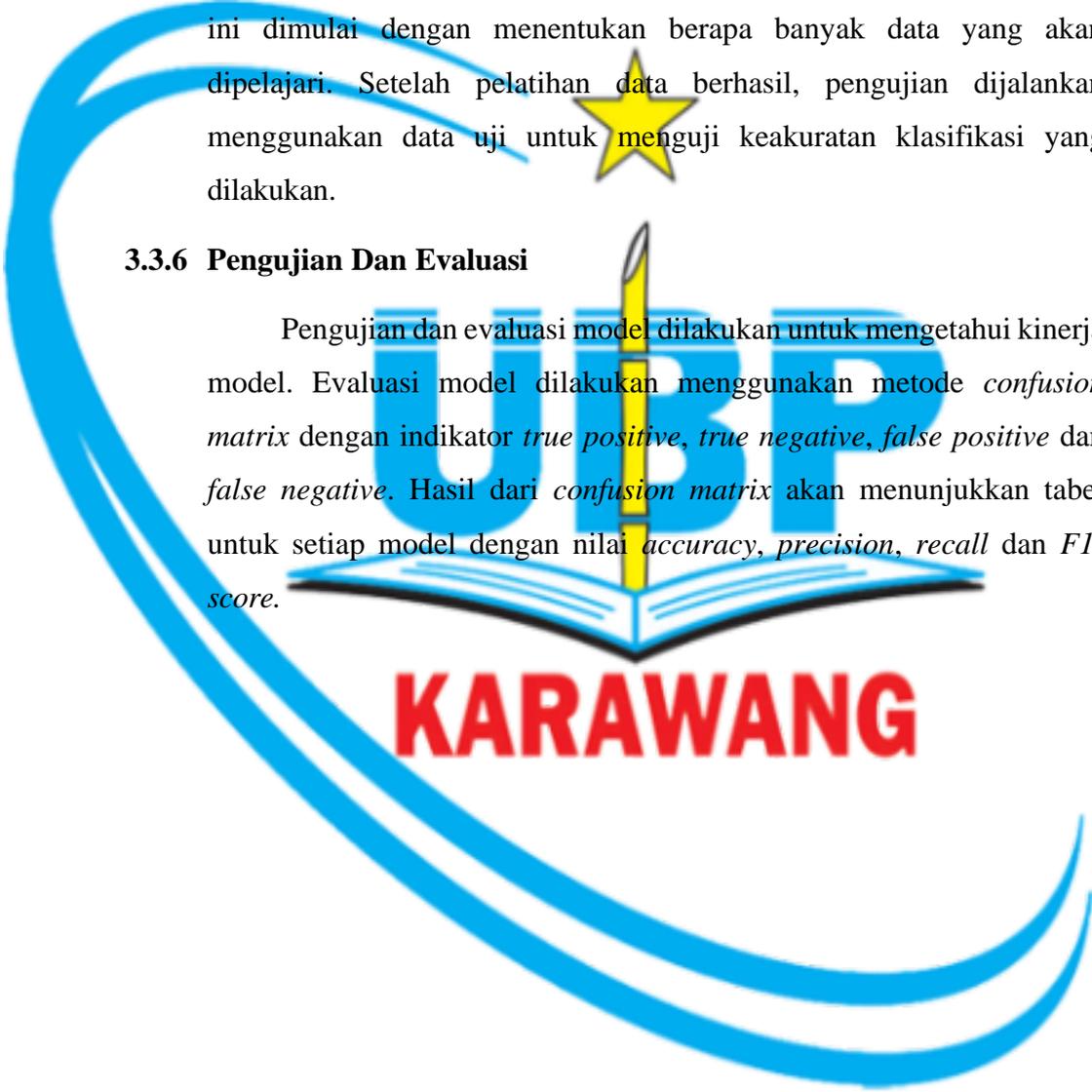
Pembobotan kata dilakukan setelah selesai melakukan tahap *labeling*. Pembobotan kata merupakan suatu metode yang digunakan untuk memberikan nilai pada frekuensi kemunculan kata dalam sebuah dokumen yang nantinya akan diproses pada tahap klasifikasi naive bayes. Semakin sering kemunculan suatu kata dalam dokumen maka akan memberikan nilai kesesuaian yang semakin besar. Dalam penelitian ini proses pembobotan kata dilakukan menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF). Setelah proses pembobotan selesai maka dokumen data telah siap untuk dilakukan training dalam tahap klasifikasi.

### 3.3.5 Klasifikasi

Algoritma yang digunakan untuk klasifikasi dalam penelitian ini adalah Naive Bayes. Dalam beberapa penelitian terkait algoritma naive bayes memiliki tingkat akurasi yang tinggi dalam proses klasifikasi. Pada proses ini data yang digunakan merupakan data bersih hasil proses *text processing* hingga pembobotan kata menggunakan TF-IDF. Proses ini dimulai dengan menentukan berapa banyak data yang akan dipelajari. Setelah pelatihan data berhasil, pengujian dijalankan menggunakan data uji untuk menguji keakuratan klasifikasi yang dilakukan.

### 3.3.6 Pengujian Dan Evaluasi

Pengujian dan evaluasi model dilakukan untuk mengetahui kinerja model. Evaluasi model dilakukan menggunakan metode *confusion matrix* dengan indikator *true positive*, *true negative*, *false positive* dan *false negative*. Hasil dari *confusion matrix* akan menunjukkan tabel untuk setiap model dengan nilai *accuracy*, *precision*, *recall* dan *F1-score*.



**KARAWANG**