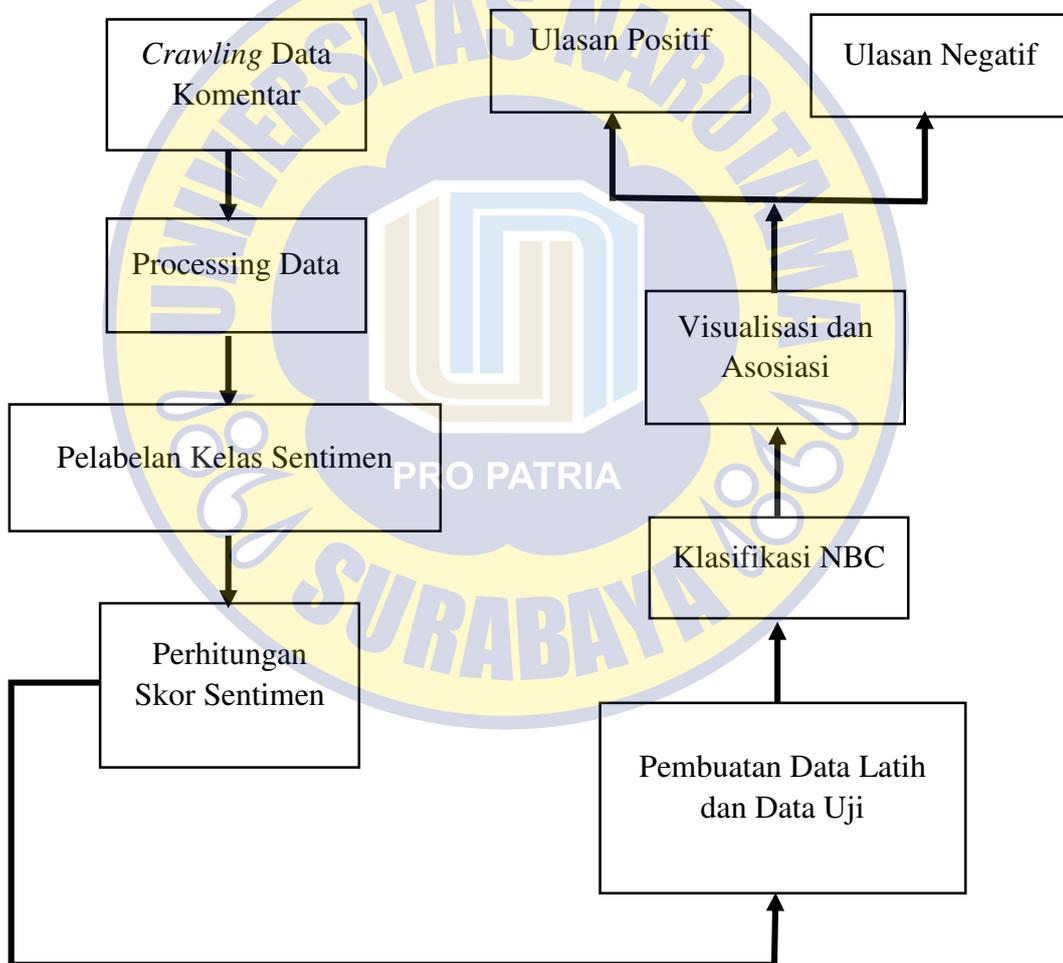


## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Perencanaan Alur Analisis Sentimen

Perencanaan alur analisis sentimen adalah sebuah gambaran umum yang menunjukkan alur dari penelitian yang sedang dilakukan pada tugas akhir ini. Alur kerja dari analisis sentimen dapat dilihat di bawah ini :



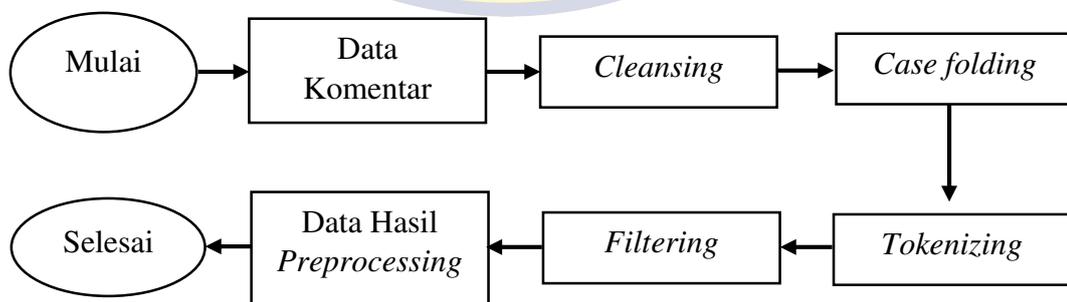
Gambar 3. 1 Diagram alir penelitian

### 3.2 Sumber Data

Dalam penelitian yang dilakukan untuk data *Crawling* atau proses pengumpulan data dilakukan dengan memanfaatkan data komentar sebagai sumber data dengan memanfaatkan *access token* yang telah disediakan oleh *Facebook*. Pada tahapan ini merupakan tahapan yang membutuhkan tingkat ketelitian serta kesabaran. Sebab data yang telah dikumpulkan harus memenuhi dari kriteria dataset yang diinginkan. Semisal dalam hal jumlah data yang diperlukan harus memenuhi target yang telah ditentukan. Selanjutnya proses pengumpulan data dilakukan sejak tanggal 17 April 2019 sampai dengan 22 Mei 2019 dari akun *Facebook* masing-masing calon presiden.

### 3.3 Preprocessing

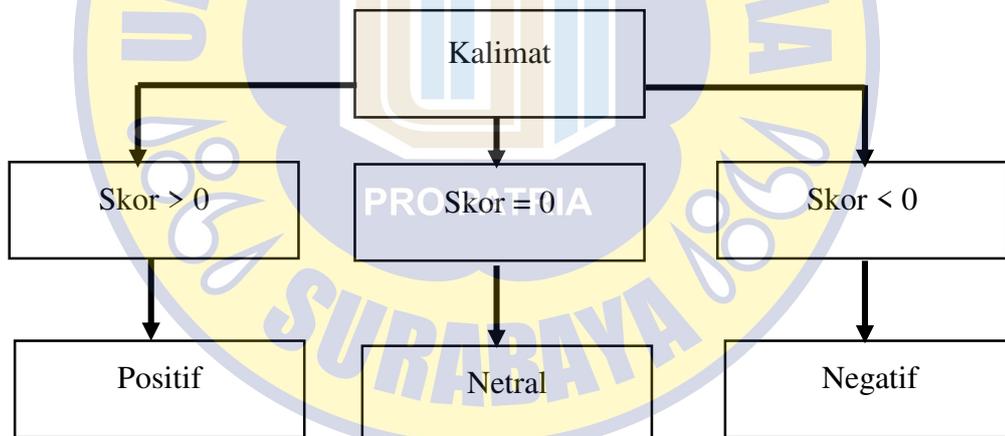
Data komentar calon presiden yang diperoleh dari proses *Crawling* pada situs *jejaring sosial Facebook* belum dapat digunakan karena memiliki bentuk kalimat atau teks tidak beraturan atau tidak terstruktur dimana memiliki banyak *noise*. Maka dari itu data tersebut perlu melakukan proses pembersihan terlebih dahulu. Adapun langkah-langkah *Preprocessing* yang akan digunakan adalah :



**Gambar 3. 2** Tahapan *Preprocessing* data

### 3.4 Pelabelan

Setelah tahapan *Preprocessing* sudah dilakukan, selanjutnya melakukan pelabelan kelas sentimen. Pada bagian ini juga merupakan salah satu proses untuk mendapatkan hasil representasi *corpus* yang diharapkan. Proses pelabelan dilakukan secara otomatis dengan cara menghitung nilai pelabelan sentimen menggunakan kamus *lexicon* dan manual dari hasil pembobotan *Term Presence*. Jika suatu kalimat memiliki skor  $> 0$  akan diklasifikasikan dalam kelas positif, jika kalimat memiliki skor  $= 0$  akan diklasifikasikan dalam kelas netral, sedangkan jika kalimat memiliki skor  $< 0$  diklasifikasikan dalam kelas negatif. Adapun proses pelabelan dapat dilihat pada gambar 3.3 berikut :



**Gambar 3. 3** Gambaran proses pelabelan kelas sentimen

### 3.5 Perhitungan

Berdasarkan teks komentar “Jalan tol yang sudah dikerjakan sangat bagus dan indah, tapi untuk pembayaran tolnya sangat mahal, terdapat 2 kata positif dan 1 kata negatif yang terdeteksi pada kamus *lexicon*, yaitu “*bagus*” dan “*indah*” sebagai kata positif, untuk kata negatif yaitu “*mahal*” sebagai kata negatif. Adapun rumus

yang digunakan dalam proses perhitungan skor sentimen adalah sebagai berikut sekaligus contoh simulasi perhitungannya.

$$\text{Skor} = (\text{Jumlah kata positif}) - (\text{Jumlah kata negatif}) \quad (3.1)$$

**Tabel 3. 1** Simulasi Perhitungan skor sentimen

Komentar	Kata Positif	Kata Negatif
Jalan tol yang sudah dikerjakan sangat bagus dan indah, tapi untuk pembayaran tolnya sangat mahal	bagus indah	mahal
<b>Jumlah</b>	<b>2</b>	<b>1</b>

Skor positif berjumlah 2 dan dikurangi skor negatif berjumlah 1, sehingga nilai akhir yang diperoleh dari simulasi perhitungan di atas mempunyai nilai 1 atau lebih dari 0, maka hasil klasifikasi komentar tersebut adalah positif.

### 3.6 Naive Bayes Classifier

Metode *Naive Bayes* (*Naive Bayes Classifier*) adalah sebuah proses klasifikasi dimana memiliki algoritma yang sangat efektif.. Data latih digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada, semakin besar data latih yang digunakan, maka akan semakin baik *machine* dalam memahami pola data. Data uji digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang digunakan untuk data latih dan data uji adalah data yang telah memiliki label kelas, dengan jumlah data latih dan data uji memiliki perbandingan 80% : 20%.

### 3.7 Visualisasi dan Asosiasi

Dalam penelitian ini dilakukan visualisasi pada semua data yang ada baik itu masuk kelas sentimen negatif maupun positif hal ini bertujuan untuk mengekstraksi informasi secara keseluruhan tentang topik atau bahasan yang sering diulas para pengguna *Facebook* terhadap calon presiden Indonesia tahun 2019. Selain itu juga dilakukan visualisasi pada masing-masing kelas sentimen untuk diambil informasi yang dianggap penting atau berguna serta dicari asosiasi antar kata pada masing-masing kelas sentimen yang paling sering muncul secara bersamaan guna memperkuat pencarian informasi tersebut.

