

BAB IV

METODOLOGI PENELITIAN

4.1 Populasi dan Sampel

Populasi dalam penelitian ini adalah *tweet* dan *retweet* dari media sosial *twitter*. Sedangkan total jumlah sampel yang digunakan dalam penelitian ini sebesar 3561 *tweet* dalam *twitter*.

4.2 Variabel Penelitian dan Definisi Operasional Variabel

Variabel penelitian yang digunakan dalam penelitian ini dapat dilihat pada tabel berikut:

Tabel 4. 1 Variabel Penelitian

Variabel	Definisi
Teks	Teks yang dimaksud merupakan tanggapan atau <i>tweet</i> dan <i>retweet</i> mengenai RKUHP
Score	Nilai klasifikasi yang telah di saring berdasarkan kamus kata positif dan negatif
Klasifikasi	Klasifikasi setiap <i>tweet</i> dibagi menjadi dua kategori yaitu sentimen positif dan negatif. Sentimen positif adalah <i>tweet</i> yang memuat opini positif. Sentimen negatif adalah <i>tweet</i> yang memuat opini negatif.

4.3 Metode Pengambilan Data

Dalam melakukan pengambilan data dari *twitter* dengan metode *scrapping* menggunakan *Twitter* API dalam kurun waktu dari tanggal 24-27 September 2019. Menggunakan kata kunci “Jokowi”, “#HidupMahasiswa”, “#SayaBersamaJokowi dan “#TrisaktiTurunLagi”. Pada penelitian ini, menggunakan metode dengan teknik *web scrapping*. Alat yang digunakan untuk *crawling* data adalah *Phyton*.

4.4 Metode Analisis Data

Dalam melakukan analisis data terdapat beberapa bantuan menggunakan *API Twitter*, *software Rstudio*, *Phyton* dan *Microsoft Excel*. Beberapa metode yang digunakan dalam penelitian ini yaitu:

1. *Text mining*

Digunakan dalam melakukan analisis data yang berupa teks yang tidak terstruktur.

2. *Word Cloud*

Word cloud digunakan untuk menampilkan hasil visualisasi data yang paling banyak digunakan.

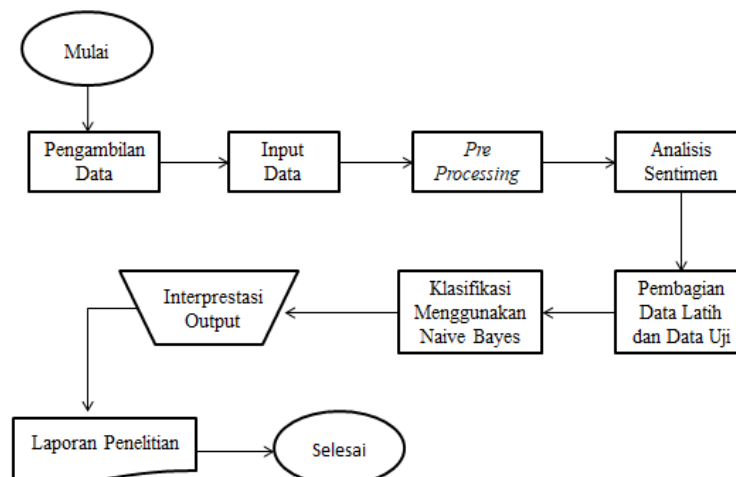
3. Analisis Sentimen

Analisis sentimen bertujuan untuk pelabelan kata-kata yang terbagi menjadi 2 yaitu positif dan negatif.

4. Klasifikasi *Naive bayes*

Setelah mendapatkan *tweet* positif dan negatif, untuk melihat tingkat akurasi yang dihasilkan menggunakan metode klasifikasi *naive bayes*.

4.5 Tahapan Penelitian



Gambar 4. 1 Tahapan Penelitian

Berikut tahapan yang penulis lakukan dalam penelitian.

1. Tahap pertama adalah pengambilan data di media sosial *twitter* dengan cara *scrapping* sesuai dengan topik yang telah ditentukan sebelumnya.

2. Kemudian, masuk kedalam tahapan penginputan data, pada tahap ini mengubah data dari *json* kedalam *csv* menggunakan *package jsonlite*.
3. Setelah data yang diperoleh berbentuk *csv* kemudian dilakukan proses *preprocessing*. Pada proses *preprocessing* terdapat tahap *cleaning*, *case folding*, *filtering*, dan *tokenizing*.
4. Setelah data diproses pada tahap *preprocessing*, tahap selanjutnya yaitu pelabelan kalimat positif dan negatif menggunakan kamus positif dan negatif untuk dilakukan analisis sentimen.
5. Tahap selanjutnya adalah pembagian data untuk memulai proses klasifikasi. Pembagian data dibagi menjadi dua yaitu pembagian data latih dan data uji, dengan jumlah data latih sebanyak 80% dan data uji sebanyak 20%.
6. Selanjutnya yaitu memasukkan data latih kedalam *script Naive Bayes* yang dilanjutkan dengan proses pengujian dengan menggunakan data uji sebesar 20% yang diambil dari data.
7. Hasil akurasi pada *naive bayes* telah didapatkan, lalu diinterpretasikan hasil dari proses kedalam laporan akhir penelitian.