

## **BAB III METODOLOGI PENELITIAN**

### **3.1 Data**

Data yang diolah untuk analisis sentimen berasal dari tweet pengguna Twitter. Data dari Twitter diambil dari hasil mention ke akun yang dituju yaitu ke akun ojek *online* (@gojekindonesia dan @GrabID),

### **3.2 Metode Pengambilan Data**

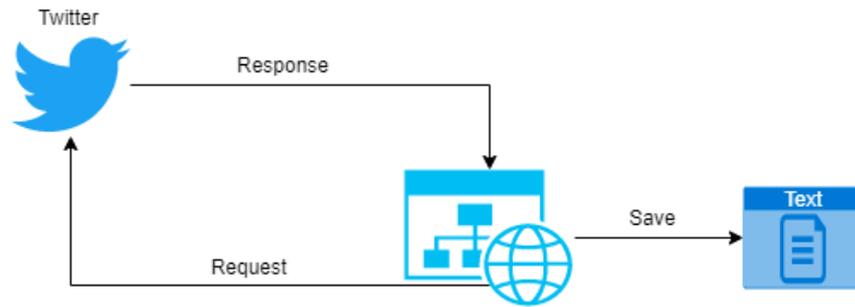
Metode pengambilan data digunakan untuk mengumpulkan data-data yang dibutuhkan dalam proses penelitian. Pengambilan data dalam penelitian ini dilakukan dengan *crawling* dan kamus jenis kata.

#### **3.2.1 Pengambilan Data untuk Kamus Jenis Kata**

Pengambilan data ini berupa *Multi Word Expression Lists*. *Multi Word Expression Lists* ini berfungsi untuk membantu dalam pengenalan jenis kata dengan kata lain kamus yang akan di gunakan untuk mendeteksi jenis kata. *Multi Word Expression Lists* diambil dari penelitian yusuf syaifudin tentang *Named Entity Recognition for Bahasa Indonesia*, yang bisa di akses pada laman <https://github.com/yusufsyaifudin/indonesia-ner>

#### **3.2.2 Pengambilan Data Komentar dengan Crawling**

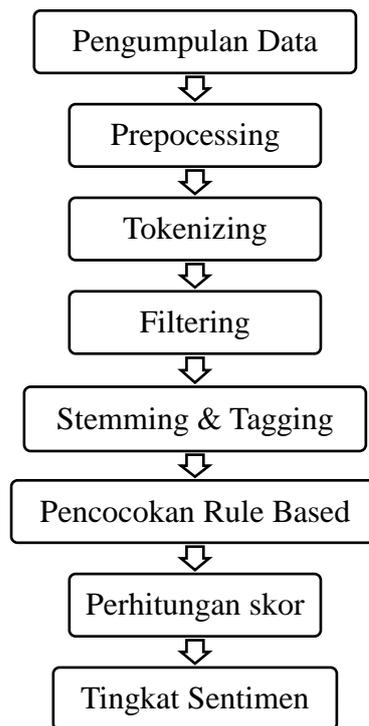
Selain itu, untuk pengumpulan dataset, diambil melalui Twitter API dengan proses *crawling*. Dataset yang ditarik dari server Twitter berupa tweet dari masyarakat mengenai respon terhadap layanan ojek online yang mencantumkan keyword akun official dari kedua ojek online yang tercantum dalam batasan masalah. Dataset yang digunakan diambil melalui proses *crawling*. *Crawling* adalah teknik untuk mengambil data dari Twitter melalui web sederhana. Dataset yang sudah di tentukan jumlah tweet dan nama akunnya akan langsung di proses sistem untuk menentukan senitemennya.



Gambar 3. 1 Pengambilan Data Komentar dengan Crawling

### 3.3 Metodologi pengolahan Data

Untuk tahap pengolahan data, dimulai dengan melakukan pengumpulan data yang didapatkan dari proses *crawling*, kemudian dilakukan tahap *preprocessing*, dilanjutkan *tokenizing* atau pemotongan *string* berdasarkan tiap kata penyusun, dan *filtering* yaitu mengambil kata penting dari hasil token. *Stemming & Tagging* untuk menghilangkan imbuhan pada kata, serta pencocokan *rule based*. Tahap akhir yaitu perhitungan skor yang dilanjutkan dengan tingkat sentimen.



Gambar 3. 2 Bagan Metodologi Penelitian

Detail dari alur pengolahan data tersebut dijelaskan pada tahapan-tahapan sebagai berikut:

- a. Tahapan pertama yakni pengumpulan data. Data diambil dari data komentar di akun resmi twitter gojek @gojekindonesia dan grab @GrabID. Pengambilan data dilakukan dengan cara *crawling*. Berikut contoh dari Data yang ada:

Tabel 3. 1 Contoh Data dari Twitter

Waktu	Komentar
2020-06-24	admin @gojekindonesia, tolong cek dm doongggg :((
2020-06-24	Eh @gojekindonesia dr sore kagak bs login ini, gak bs balik Knp gak ada opsi OTP via call sih? Nyebelin sumpah
2020-06-24	Kaget bgt @gojekindonesia php 400+ karyawannya :( Padahal aku baru aja browsing buat apply job di sana :(

- b. Hasil *crawling* disimpan di dalam database, untuk selanjutnya data akan otomatis menampilkan sentiment hasil dari sistem.
- c. Untuk tahap selanjutnya yaitu *text preprocessing*. Dimana proses klasifikasi pada tahapan ini akan dilakukan pengecekan apakah data mengandung Case Folding, tanda baca, penghapusan dan konversi nomor, penghapus pengulangan dan singkatan juga terjemahan Bahasa non formal. Hal tersebut untuk mengurangi bagian tidak penting pada proses analisis sentiment. Contoh proses *text preprocessing* adalah

Tabel 3. 2 Contoh hasil text preprocessing

Tweet	Preprocessing
admin @gojekindonesia, tolong cek dm doongggg :((	Admin tolong cek dm doongggg
Eh @gojekindonesia dr sore kagak bs login ini, gak bs balik Knp gak ada opsi OTP via call sih? Nyebelin sumpah	Eh dr sore kagak login ini, gak bs balik knp gak ada opsi otp via call sih? Nyebelin sumpah
Kaget bgt @gojekindonesia php 400+ karyawannya :( Padahal aku baru aja browsing buat apply job di sana :(	kaget bgt phk karyawannya padahal aku baru aja browsing buat apply job di sana

d. Setelah proses text preprocessing selesai, proses selanjutnya yakni text mining. Tahapan text mining adalah

- *Tokenizing* merupakan proses untuk memecah urutan karakter menjadi kata per kata
- *Filtering* merupakan proses menghapus kata tidak penting berdasarkan stopwords yang sudah ada.
- *Stemming* merupakan mencari kata dasar dan menghilangkan imbuhan seperti: membayar menjadi bayar. *Stemming* ini menggunakan *library* dari sastrawi
- *Tagging* merupakan pelabelan kata berdasarkan kamus yang sudah ada. Pada proses *tagging* ini jika menemui 1 kata dalam 2 jenis kata, maka sistem akan mengambil jenis kata berdasarkan urutan jenis kata pada kamus.

Tabel 3. 3 Contoh hasil text mining

<b>Preprocessing</b>	<b>Hasil Text Mining</b>
admin tolong cek dm doongggg	tolong cek dm doongggg
Eh dr sore kagak login ini, gak bs balik knp gak ada opsi otp via call sih? Nyebelin sumpah	Eh dr sore kagak login ini, gak bs balik knp gak ada opsi otp via call sih? Nyebelin sumpah
kaget bgt phk karyawannya padahal aku baru aja browsing buat apply job di sana	kaget bgt phk karyawannya padahal browsing apply job

e. Tahap selanjutnya yaitu analisis sentiment dengan rule based yang sudah di jelaskan di atas. Kata yang sudah melalui tahapan preprocessing dan text mining akan di klasifikasikan berdasarkan rule based yang ada dan tahapan word degree ini, kata akan di klasifikasikan apakah kata tersebut bersentimen positif, negative ataupun netral. Setiap kata memiliki nilai yang berbeda. Maka proses pemberian nilai sebagai berikut:

Tabel 3. 4 Contoh hasil analisis sentimen

<b>Tweet</b>	<b>Sentiment</b>
tolong cek dm doongggg	tolong Verba 1 cek Nomina 0 <b>= 1</b>
Eh dr sore kagak login ini, gak bs balik knp gak ada opsi otp via call sih? Nyebelin sumpah	sore Nomina 0 kagak Adverbia 0 opsi Nomina 0 via Preposisi 0 sumpah Nomina 0 <b>= 0</b>
kaget bgt phk karyawannya padahal browsing apply job	kaget Adjektiva -1 karyawan Nomina 0 <b>= -1</b>

f. Akurasi

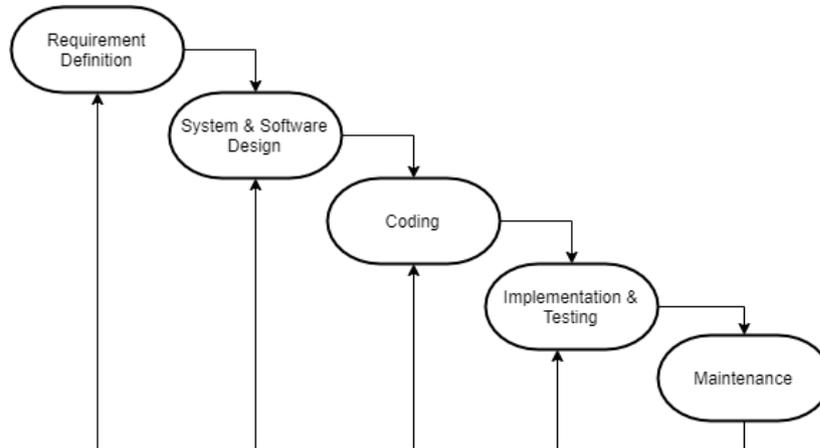
Akurasi diperoleh dengan membandingkan jumlah data hasil klasifikasi (prediksi) yang sesuai dengan jumlah keseluruhan data. Semakin tinggi nilai akurasi yang diperoleh, maka hasil klasifikasi semakin baik.

Tabel 3. 5 Contoh perbandingan dengan manual opinion

<b>Tweet</b>	<b>Sentiment</b>	<b>Manual</b>	<b>akurasi</b>
admin @gojekindonesia, tolong cek dm doongggg :(	1	-1	salah
Eh @gojekindonesia dr sore kagak bs login ini, gak bs balik Knp gak ada opsi OTP via call sih? Nyebelin sumpah	0	-1	salah
Kaget bgt @gojekindonesia php 400+ karyawannya :( Padahal aku baru aja browsing buat apply job di sana :(	-1	-1	benar

### 3.4 Metode Pengembangan Perangkat Lunak

Metode yang digunakan dalam pengembangan sistem ini adalah metode waterfall. Metode *Waterfall* merupakan model pengembangan sistem informasi yang sistematis dan sekuensial (Sasmito, 2017).



Gambar 3. 3 Metode Waterfall

Metode *Waterfall* memiliki tahapan-tahapan sebagai berikut:

a. *Requirements Definition*

*Requirement Definition* merupakan langkah dalam mendefinisikan kebutuhan yang diperlukan dalam penelitian seperti kebutuhan data *system*. Kebutuhan data *system* didapatkan melalui beberapa langkah yaitu studi literatur dan observasi.

b. *System and Software Design*

Desain Sistem adalah persiapan rancang bangun implementasi yang menggambarkan bagaimana suatu sistem dibentuk yang berupa penggambaran, perencanaan dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah kedalam satu kesatuan yang utuh dan berfungsi, menyangkut di dalamnya konfigurasi komponen perangkat lunak dan perangkat keras dari suatu sistem.

c. *Coding*

*Coding* merupakan tahap untuk menerjemahkan hasil proses perancangan menjadi sebuah bentuk program komputer yang dimengerti oleh mesin komputer.

d. *Implementation and Testing*

Pengujian dilakukan berdasarkan tujuan dari rumusan masalah. Pengujian dilakukan dengan cara membangun sebuah prototype aplikasi analisis sentimen dan pengujian dengan mencoba sistem diuji untuk mengecek setiap kegagalan maupun kesalahan.

e. *Maintenance*

Pemeliharaan akan dilakukan apabila ada update fitur atau memperbaiki kesalahan yang ditemukan pada saat sistem digunakan.

### **3.5 Metode Pengujian**

#### **3.5.1 Pengujian Fungsional/ Validasi**

Pengujian fungsional menggunakan Black Box Testing yaitu pengujian yang didasarkan pada detail aplikasi seperti tampilan aplikasi, fungsifungsi yang ada pada aplikasi, dan kesesuaian alur fungsi dengan bisnis proses yang diinginkan oleh customer. Pengujian ini tidak melihat dan menguji source code program

#### **3.5.2 Pengujian Akurasi Metode**

Pengukuran kinerja dari sistem / metode klasifikasi dilakukan dengan cara membandingkan sentimen manual dengan sentimen sistem, sentimen manual dilakukan dengan bantuan 7 orang. Pengujian dilakukan dengan menghitung nilai *accuracy*, *recall* dan *precision*. *Accuracy* diperoleh dengan membandingkan jumlah data hasil klasifikasi (prediksi) yang sesuai dengan jumlah keseluruhan data. Semakin tinggi nilai akurasi yang diperoleh, maka hasil klasifikasi semakin baik. Akan tetapi, jika hanya melihat nilai akurasi saja tidak bisa mendeteksi adanya penyimpangan data. Oleh karena itu kami juga menghitung nilai *recall* dan *precision*. *Recall* diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total data yang dianggap relevan. Sedangkan *precision* diperoleh dengan membandingkan jumlah data hasil klasifikasi yang relevan dan total jumlah data yang ditemukan pada kelas tertentu. (Lailiyah, 2017)

Rumus untuk menghitung nilai *accuracy* adalah

$$Akurasi = \frac{jumlah\ sentimen\ benar}{jumlah\ data} * 100\% \quad (1)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (2)$$

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (3)$$

F-measure merupakan parameter tunggal ukuran keberhasilan retrieval yang menggabungkan recall dan precision. Nilai F-measure didapat dari perhitungan hasil perkalian precision dan recall dibagi dengan hasil penjumlahan precision dan recall kemudian dikalikan dua dan perhitungan f-measure menggunakan Persamaan (4).

$$f - measure = 2 * \frac{precision * recall}{precision + recall} \quad (4)$$

### 3.5.3 K-fold Cross Validation

K-fold adalah salah satu metode Cross Validation yang populer dengan melipat data sebanyak K dan mengulangi (iterasi) eksperimennya sebanyak K juga. Pada pengujian ini data yang akan digunakan adalah 1000 tweet termasuk data uji yang nantinya akan dibagi menjadi 10 bagian atau k=10 sehingga data yang diperoleh adalah 1000 data dibagi menjadi 10 lipatan. Selain itu akan ditentukan mana yang termasuk data training dan mana yang termasuk data testing dengan perbandingan 8:2. Pengujian menggunakan data yang sudah dipartisi akan diulang sebanyak 10 kali (K=10) dengan posisi data testing berbeda disetiap iterasinya. Misalkan iterasi pertama data tes pada posisi awal, iterasi kedua data testing di posisi kedua begitu seterusnya(Rahmawati,2018)

Total Sampel = 1000									
iterasi 1/10	Test set								
iterasi 2/10		Test set							
iterasi 3/10			Test set						
iterasi 4/10				Test set					
iterasi 5/10					Test set				
iterasi 6/10						Test set			
iterasi 7/10							Test set		
iterasi 8/10								Test set	
iterasi 9/10									Test set
iterasi 10/10									Test set

Gambar 3.4 Ilustrasi 10-fold Cross Validation