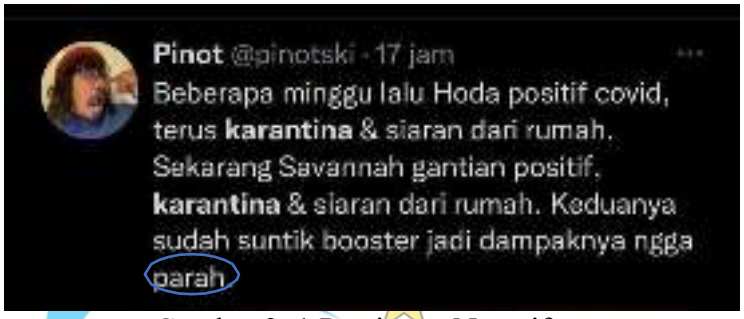
BAB III
METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian dilakukan pada media sosial *twitter* dengan mengambil sampel dari postingan yang *tweet,* oleh masyarakat ditunjukan pada gambar 3.1 postingan positif dan gambar 3.2 postingan negatif.



Gambar 3. 1 Postingan Negatif

Gambar 3. 2 Postingan Positif

Dari gambar yang di tunjukan oleh 3.1 terdapat kalimat "Keduanya sudah suntik booster jadi dampaknya ngga parah" terdapat kata negatif pada kata "Parah" Gambar 3.2 terdapat*sentimen* positif pada kalimat "Karantina hanya berlaku untuk anggota NKRI" terdapat kata positif pada kata "Berlaku".

3.2. Lokasi dan Waktu Penelitian

Lokasi penelitian ini dilakukan di Laboratium Riset Universitas Buana Perjuangan Karawang sejak bulan November sampai Desember.

Tabel 3. 1 Tabel Kegiatan

	November				Desember				Januari				Februari			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Studi literatur																
Pengumpulan data																
Text preprocessing																

	Maret				April				Mei			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4
Pengujian												
Evaluasi												

Gambar 3. 3 Prosedur Penelitian

3.2.1. Analisis Data

Analisis data dilakukan untuk mengetahui permsalahan pada media *twitter* yaitu opini masyrakat tentang Karantina. Analisis data dilakukan untuk memperoleh sentimen berupa positif, negatif dan netral.

3.2.2. Pengumpulan Data

Penelitian ini dimulai dengan pengumpulan data melalui media sosial yaitu *twitter.* Pengumpulan data untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Data diperoleh sejak 04 Januari – 5 Januari 2022 dengan lebih dari 100 data.

3.2.3. Text Pocessing

Dalam tahapan *text* preprocessing suatu proses pengubahan bentuk data yang tidak terstruktur menjadi data yang terstruktur untuk sesuai informasi yang dibutuhkan.

Berikut tahapan-tahapan *text preprocessing*:

1. Case Folding

   Merupakan prosese semua huruf dalam dokumen kecil, hanya huruf ‘a’ sampaidengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggapdelimeter.

2. Tokenzing

   Tokenzing merupakan tahapan penguraian string teks menjadi trmatau kata. Tujuan dari *Tokenization* yaitu memisahkan kata-kata dalam sebuah paragraf, kalimat atau halaman ke dalam kata tunggal.

3. Filtering

   Filtering pada tahap ini mengambil kata-kata penting yang telah dipotong kata per kata dilakukan *filter*. Dengan dilakukanya pembuangan kata-kata yangdianggap tidak penting seperti tanda hubung dan lainya.

4. Stemming

   Stemming merupakan tahapan pengubahan suatu bentuk kata menjadi akar katanya dengan menghilangkan imbuhan awal dan akhir pada kata tersebut.

3.2.4. Algoritma K-Means

Algoritma *K-Means* merupkan salah satu metode pengelompokan data nonhierarki (sekatan)yang berusaha mempartisi data yang ada ke dalam bentuk atau lebih. Dalam tahap ini untuk

pengelompokan sentimen berupa positf, negatif dan netral.

Berikut tahapan-tahapan *K-Means*:

1. Tentukan jumlah klaster yang inigin dibentuk dan tetapkan pusat klaster k.
2. Menggunakan jarak *euclidean* kemudian hitung setiap data ke pusat ke klaster.

(4)

$$d(i,k)\sqrt{\sum_{i}^{m}(C_ij - C_kj)^2}$$

3. Kelompokkan data ke dalam klaster dengan jarak yang paling pendek dengan persamaan.

$$min\sum_{k}^{i} - a_{ik} - = \sqrt{\sum_{i}^{m}(C_ij - C_kj)^2} \quad (5)$$

4. Hitung pusat klaster yang baru menggunakan persamaan.

$$C_{kj} = \frac{\sum_{k}^{i} x_{ij}}{p} \quad (6)$$

Dengan : X ij Kluster ke k p = banyaknya anggota kluster ke – k.

5. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindahke kluster yang lain.

3.2.5. Pengujian

Dalam tahap pengujian ini agar semua fungsi bekerja dengan baik atau tidak, tujuan dilakukanya pengujian ini untuk mengetahui tingkat akurasi melalui metode *Sum Square of Error* (SSE).

Varian = SSE/n, proses menghitung varian untuk seluruh populasi.

1. Varian = SSE/(n-1), proses menghitung varian sample.

Varian = $\frac{SEE}{(n-1)}$ 7)

Varian = $\frac{6{,}921}{(9)}$ 8)

Varian = $0{,}769$ 9)