

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian pada media sosial *twitter* dengan mengambil sampel dari postingan yang di *tweet* oleh masyarakat ditunjukkan pada Gambar 3.1 postingan positif dan Gambar 3.2 postingan negatif



Gambar 3.1 Postingan Positif



Gambar 3.2 Postingan Negatif

Berdasarkan yang ditunjukkan oleh Gambar 3.1 terdapat kalimat “Lurah Selat Selatan mendukung kebijakan PPKM dari pemerintah guna memutus mata rantai omicron” mengandung makna positif pada kata “Memutus” Gambar 3.2 terdapat sentimen negatif pada kalimat “Dampak Negatif PPKM bagi pedagang luar biasa semua resah, kesal dan stress” pada kata “Kesal”.

3.2. Lokasi dan Waktu Penelitian

Lokasi penelitian ini dilakukan di Laboratorium Riset Universitas Buana Perjuangan Karawang. Waktu penelitian dilakukan sejak bulan November sampai dengan Mei. Perincian penelitian ditunjukkan pada Tabel 3.1

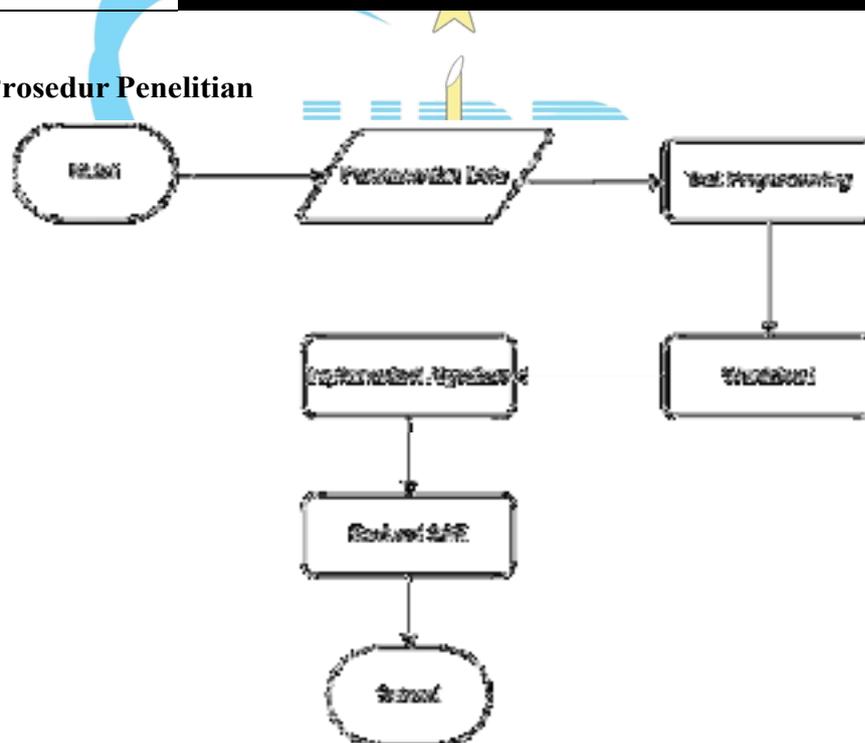
Tabel 3.1 Waktu Penelitian

	November				Desember				Januari				Februari			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Studi Literatur	[Redacted]															
Proposal	[Redacted]															

	November				Desember				Januari				Februari			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pengumpulan data	█															
Text preprocessing					█											
Analisis sentimen									█							

	Maret				April				Mei			
Kegiatan	1	2	3	4	1	2	3	4	1	2	3	4
Klasifikasi	█											
Evaluasi					█				█			

3.3. Prosedur Penelitian



Gambar 3.3 Prosedur Penelitian

3.3.1. Analisis Data

Analisis data dilakukan untuk mengetahui permasalahan yang terdapat pada media sosial *twitter* yaitu opini masyarakat tentang PPKM. Analisis data dilakukan untuk menemukan dan memperoleh sentimen positif, negatif dan netral

3.3.2. Pengumpulan Data

Penelitian ini dimulai dengan melakukan pengumpulan data melalui salah satu media sosial *twitter*. Pengumpulan data untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Data diperoleh sejak 05 januari 2022 dengan lebih dari 1200 data.

3.3.3. Text Preprocessing

Pada tahap *text preprocessing*, informasi yang telah diperoleh akan dilakukan proses penyeleksian data dengan tujuan menjadikan data tidak terstruktur menjadi data terstruktur. Sebelum dilakukannya proses klasifikasi, teks dokumen atau data harus disiapkan terlebih dahulu, dan proses tersebut dinamakan *Text Preprocessing*. Didalam tahapan *text preprocessing* suatu proses perubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan. Terdapat beberapa tahapan pada *Text Preprocessing* yaitu *Case Folding*, *Tokenizing*, *Filtering* dan TF IDF sebagai berikut (Yudiarta et al., 2018) :

1. Case Folding

adalah mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimeter

2. Tokenizing

Tahapan ini adalah tahap pemotongan string input berdasarkan tiap kata yang menyusun untuk menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan Panjang teks. Dengan dilakukannya perubahan karakter huruf menjadi huruf kecil yang disebut dengan normalisasi kata. Proses *tokenizing* yang dilakukan yaitu menghilangkan delimeter-delimeter yaitu symbol dan tanda baca yang ada pada teks tersebut seperti \$, @, &, tanda koma (,), tanda titik (.), tanda tanya (?), tanda seru (!).

3. Filtering

Pada tahap ini mengambil kata-kata penting yang telah dipotong kata per kata dilakukan *filter*. Tahap dilakukan menggunakan algoritma *stoplist* (membuang kata kurang penting) dan *wordlist* (menyimpan kata penting). Beberapa contoh *Stopword* yaitu “yang”, “dan”, “di” dan lain

lain. *Stoplist* atau *stopword* kata yang termasuk ke dalam *stoplist* akan dibuang dan tidak digunakan pada proses selanjutnya. Sedangkan *wordlist* kata yang termasuk kedalam *stoplist* akan digunakan pada proses selanjutnya.

4. TF-IDF

TF-IDF tahapan terakhir dalam proses *text processing* untuk melakukan pembobotan kata. TF-IDF sebuah statistic numerik yang dapat menunjukkan relevansi kata kunci dengan kata tertentu. Selain itu, TF IDF juga dapat mengetahui kata apa yang sering muncul pada suatu dokumen.

3.3.4. Algoritma DBSCAN

Algoritma DBSCAN metode *clustering* yang berbasis kepadatan (*density-based*) dengan prinsip pengelompokan data untuk mencari kata tertentu dari sebuah dokumen. DBSCAN berguna untuk melakukan *clustering* pada data yang banyak mengandung *noise*, Dikarenakan DBSCAN tidak akan memasukkan data yang dianggap *noise* kedalam *cluster* manapun. DBSCAN untuk setiap titik dari sebuah *cluster*, *neighborhood* dari radius yang diberikan harus mengandung setidaknya jumlah minimum poin, yaitu kepadatan *neighborhood* harus melebihi beberapa *threshold* ditetapkan. Algoritma ini membutuhkan dua parameter masukan untuk bisa melakukan *clustering* yaitu (goleman, daniel; boyatzis, Richard; Mckee, 2019).

3.3.5. Evaluasi

Dalam tahap pengujian yaitu untuk mengetahui semua fungsi berkerja dengan baik atau tidak, tujuan dilakukannya pengujian ini untuk mengetahui tingkat akurasi menggunakan *Sum Square of Error* (SSE) dari hasil prediksi yang akan diperoleh. (Refialy et al., 2021)

1. Varian = SSE/n , proses menghitung varian untuk seluruh populasi.
2. Varian = $SSE/(n-1)$, proses menghitung varian sample

Varian	=	$\frac{SSE}{(n - 1)}$	1)
Varian	=	$\frac{6,921}{(9)}$	2)
Varian	=	0,769	3)

