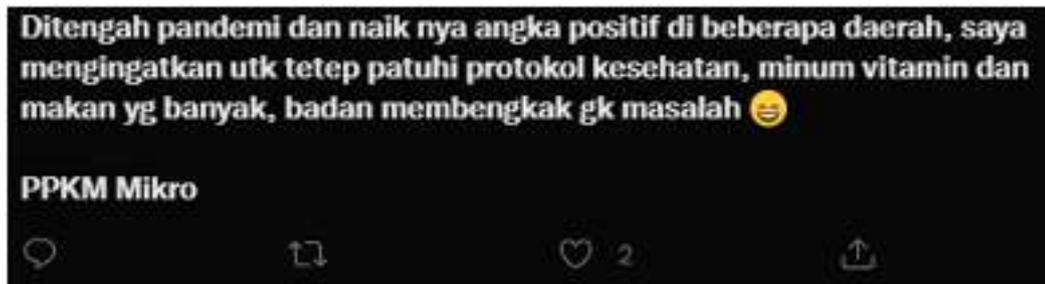


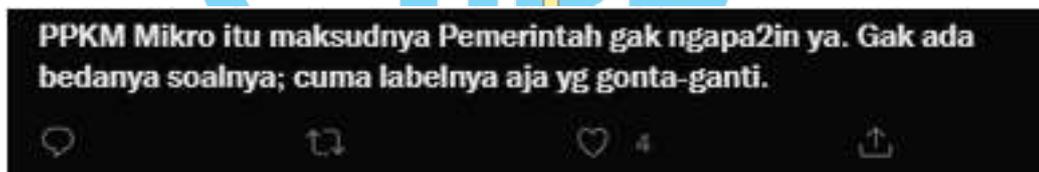
BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian yang digunakan pada penelitian ini adalah Tweet dari para pengguna Twitter yang berkaitan dengan kata PPKM yang diambil pada saat terjadinya PPKM Mikro, yaitu pada tanggal 11 Maret 2021 sampai 23 Juni 2021 sebanyak 50.000 Tweet. Objek tersebut nantinya akan diklasifikasikan ke dalam sentimen positif dan negatif.



Gambar 3.1 Tweet Positif



Gambar 3.2 Tweet Negatif

Gambar di atas merupakan sampel Tweet sentimen positif dan negatif dari pengguna Twitter mengenai PPKM. Tweet dengan sentimen positif ditunjukkan oleh gambar 3.1 yang terdapat pada kalimat “tetep patuhi protokol kesehatan”. Sementara sentimen negatif ditunjukkan oleh Gambar 3.2 pada kalimat “PPKM Mikro itu maksudnya Pemerintah gak ngapa2in ya”.

3.1.1 Peralatan Penelitian

Penelitian ini membutuhkan peralatan, yaitu perangkat keras dan perangkat lunak. Spesifikasi perangkat keras terdiri dari *Processor Core i5-6200U*, 8192MB RAM, dan 1TB HDD. Perangkat lunak yang dimanfaatkan untuk pengolahan bahan penelitian adalah sistem operasi Windows 11, bahasa pemrograman Python, dan Visual Studio Code.

3.1.2 Lokasi dan Waktu Penelitian

Lokasi penelitian dilakukan di kampus Universitas Buana Perjuangan Karawang yang dimulai pada bulan Oktober 2021. Perincian waktu dapat dilihat pada tabel 3.1.

Tabel 3.1 Jadwal Penelitian

Kegiatan	Bulan 1				Bulan 2				Bulan 3				Bulan 4			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Studi Literatur																
Pengumpulan Data																
Kegiatan	Bulan 5				Bulan 6											
	1	2	3	4	1	2	3	4								
Analisis Data																
<i>Text Preprocessing</i>																
Pembobotan Kata																
Proses Klasifikasi																
Evaluasi																

3.2 Prosedur Penelitian

Prosedur penelitian digambarkan dalam bentuk diagram alir pada Gambar 3.3. Penelitian dimulai dari pengumpulan data. Selanjutnya, data tersebut dilakukan *preprocessing* agar terstruktur dan menghilangkan *noise* atau derau. Setelah itu dilakukan pembobotan kata dengan TF-IDF. Lalu, proses klasifikasi menggunakan algoritma SVM dan terakhir proses evaluasi dengan menggunakan *Confusion Matrix*.



Gambar 3.3 Prosedur Penelitian

Prosedur penelitian dilakukan secara bertahap, dimulai dari pengumpulan data, *text preprocessing*, pembobotan kata, proses klasifikasi, dan evaluasi. Pengumpulan data diambil dari media sosial Twitter yang mengandung kata PPKM. Data yang telah diambil akan masuk pada tahap *text preprocessing*, yaitu proses pembersihan data. Proses pembobotan kata, yaitu proses pengubahan data teks menjadi data numerik. Proses selanjutnya merupakan proses klasifikasi menggunakan SVM dan proses evaluasi untuk mengukur tingkat akurasi.

3.2.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara mengumpulkan Tweet berbahasa Indonesia pada media sosial Twitter yang membahas mengenai PPKM. Data diambil pada tanggal 11 Maret 2021 sampai 23 Juni 2021 tanggal tersebut dipilih karena bertepatan dengan PPKM Mikro.

3.2.2 Text Preprocessing

Data yang sudah dikumpulkan belum bisa langsung digunakan untuk klasifikasi dikarenakan data tersebut masih kotor. Tahapan *Text Preprocessing* bertujuan untuk membersihkan data serta mengurangi volume data. Terdapat beberapa tahapan pada *Text Preprocessing*, sebagai berikut:

a. Penghapusan Data

Tweet yang memiliki kalimat yang sama akan dihapus sehingga hanya Tweet sentimen positif dan negatif yang masuk pada langkah berikutnya, proses akan lebih cepat karena data sudah berkurang.

b. Case Folding

Case folding merupakan proses untuk mengubah huruf menjadi kecil. Proses ini dilakukan karena tidak semua *Tweet* yang ada pada data memiliki huruf kecil semua.

Tabel 3.2 Contoh *Case Folding*

Sebelum	Sesudah
@CNNIndonesia sebaiknya pahami definisi dn konsekwensi lockdown dn PPKM biar tdk rancu.	@cnnindonesia sebaiknya pahami definisi dn konsekwensi lockdown dn ppkm biar tdk rancu.

utamakan nyawa rkyt, jgn utamakan konglomerat..	utamakan nyawa rkyt, jgn utamakan konglomerat..
---	---

c. *Cleansing*

Proses *cleansing*, yaitu proses pembersihan pada *Tweet* seperti tanda baca, angka, URL, tagar, emotikon, dan *mention*. Proses ini dilakukan agar *Tweet* memiliki huruf saja.

Tabel 3.3 Contoh *Cleansing*

Sebelum	Sesudah
bukan lockdown, pemerintah pilih penguatan ppkm mikro mulai 22 juni hingga 5 juli 2021 https://t.co/qrffjubzwmj	bukan lockdown pemerintah pilih penguatan ppkm mikro mulai juni hingga juli

d. Normalisasi Kata

Normalisasi kata merupakan proses perubahan kata-kata yang tidak baku seperti singkatan dan bahasa gaul menjadi baku. Contohnya seperti kata “knp” menjadi kata “kenapa”, kata “gpp” menjadi “tidak apa-apa”, dan yang lainnya.

Tabel 3.4 Contoh Normalisasi Kata

Sebelum	Sesudah
sebaiknya pahami definisi dn konsekwensi lockdown dn ppkm biar tdk rancu utamakan nyawa rkyt jgn utamakan konglomerat	sebaiknya pahami definisi dan konsekuensi lockdown dan ppkm biar tidak rancu utamakan nyawa rakyat jangan utamakan konglomerat

e. *Tokenizing*

Proses ini merupakan proses pemisahan kata pada kalimat. Tujuan dari proses ini adalah untuk mempermudah proses pada tahap selanjutnya. Umumnya setiap kata akan dipisahkan dengan karakter spasi.

Tabel 3.5 Contoh *Tokenizing*

Sebelum	Sesudah
sebaiknya pahami definisi dan konsekuensi lockdown dan ppkm biar tidak rancu utamakan nyawa rakyat jangan utamakan konglomerat	[sebaiknya, pahami, definisi, dan, konsekuensi, lockdown, dan, ppkm, biar, tidak, rancu, utamakan, nyawa, rakyat, jangan, utamakan, konglomerat]

f. *Stopword Removal*

Tahap ini merupakan tahap selanjutnya setelah proses *tokenizing* selesai. Tahapan ini akan mengambil kata penting dari hasil *tokenizing* dan menghilangkan *stopwords*. *Stopwords* adalah kata-kata yang tidak deskriptif dan bukan kata penting. Beberapa contoh kata-kata pada *stopwords*, yaitu “yang”, “di”, “ini”, “sudah”, dan yang lainnya.

Tabel 3.6 Contoh *Stopword Removal*

Sebelum	Sesudah
[kepala, negara, presiden, joko, widodo, memandang, kebijakan, ppkm, mikro, menjadi, kebijakan, pengendalian, covid, covid, yang, paling, tepat, dalam, situasi, saat, ini, karena, dinilai, bisa, mengendalikan, covid, tanpa, mematikan, ekonomi, rakyat]	[kepala, negara, presiden, joko, widodo, memandang, kebijakan, ppkm, mikro, kebijakan, pengendalian, covid, situasi, dinilai, mengendalikan, covid, mematikan, ekonomi, rakyat]

g. *Lemmatizing*

Lemmatizing merupakan sebuah proses mengubah suatu kata menjadi kata dasar dengan mengetahui isi dari kata tersebut. Berbeda dengan *stemming* yang beroperasi pada satu kata tanpa mengetahui isi dari kata tersebut. Contoh kalimat “orang itu membelikannya sebuah makanan untuk ikannya” dengan menggunakan *stemming* akan menghasilkan “orang itu beli buah makan untuk ikan” sedangkan

dengan *lemmatizing* akan menghasilkan “orang itu beli sebuah makan untuk ikannya.

Tabel 3.7 Contoh *Lemmatizing*

Sebelum	Sesudah
[pemerintah, tindakan, cepat, penguatan, ppkm, mikro, mendorong, percepatan, pelaksanaan, vaksinasi]	['perintah', 'tindak', 'cepat', 'kuat', 'ppkm', 'mikro', 'dorong', 'cepat', 'laksana', 'vaksinasi']

3.2.3 Pembobotan Kata

Tahap selanjutnya adalah menghitung bobot dari setiap kata berdasarkan banyaknya kata yang sering muncul pada dokumen. Pembobotan kata ini juga mengubah data teks menjadi data numerik agar nantinya bisa diproses oleh algoritma. Metode yang digunakan pada tahapan ini adalah TF-IDF. TF-IDF merupakan statistik numerik yang dapat menunjukkan kata kunci dengan kata tertentu. TF-IDF juga dapat mengetahui kata apa saja yang sering muncul. *Term Frequency-Invers Document Frequency* (TF-IDF) adalah proses perubahan data dari data tekstual menjadi data numerik. TF-IDF memiliki dua perhitungan *Term Frequency* (TF) dan *Invers Document Frequency* (IDF).

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

Keterangan:

$f_{t,d}$: jumlah *term* dari dokumen.

t : berapa kali *term* yang muncul pada dokumen.

d : dokumen

$$idf(t) = \log \log \frac{1 + n}{1 + df(t)} + 1 \quad (2)$$

Keterangan:

n : jumlah dokumen dalam kumpulan dokumen.

$df(t)$: jumlah dokumen dalam kumpulan dokumen yang berisi *term*.

t : term

Hasil dari TF dan IDF bisa mendapatkan bobot kata yang disebut TF-IDF dengan persamaan.

$$tf - idf(t, d) = tf(t, d) \times idf(t) \quad (3)$$

3.2.4 Proses Klasifikasi

Pada tahapan ini akan menggunakan algoritma SVM untuk melakukan proses klasifikasi berdasarkan bobot nilai yang telah dilakukan sebelumnya. Algoritma SVM berguna untuk melakukan klasifikasi pada suatu data pembelajaran. Ada beberapa tahapan algoritma SVM, sebagai berikut:

1. Memasukkan data ke ruang dimensi.
2. Mencari margin terbaik pada *support vector*.
3. Menentukan *hyperplane*.
4. Melakukan klasifikasi dengan menggunakan *kernel* linear.

Pada penelitian ini akan mengimplementasikan klasifikasi algoritma SVM dengan penyelesaian dalam bahasa pemrograman Python.

3.2.5 Evaluasi

Pengujian pada penelitian ini untuk mengevaluasi hasil dari algoritma yang digunakan. Metode evaluasi yang digunakan *Confusion Matrix*. Metode ini cukup membantu untuk proses analisis kualitas *classifier*. Pengujian akan menghitung nilai *accuracy*, *recall*, dan *precision* yang akan ditampilkan dalam bentuk persentase.