

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek pada penelitian ini sentiment atau opini masyarakat terhadap Vaksin Covid-19 yang diambil pada media sosial *Twitter*. Penelitian dilakukan di Universitas Buana Perjuangan Karawang berlangsung pada bulan Oktober 2021.

Tabel 3. 1 Jadwal Penelitian

	Bulan							
Kegiatan	1	2	3	4	5	6	7	8
Pengumpulan Data								
Labelling								
Text Pre-processing								
Klasifikasi								
Pengujian								

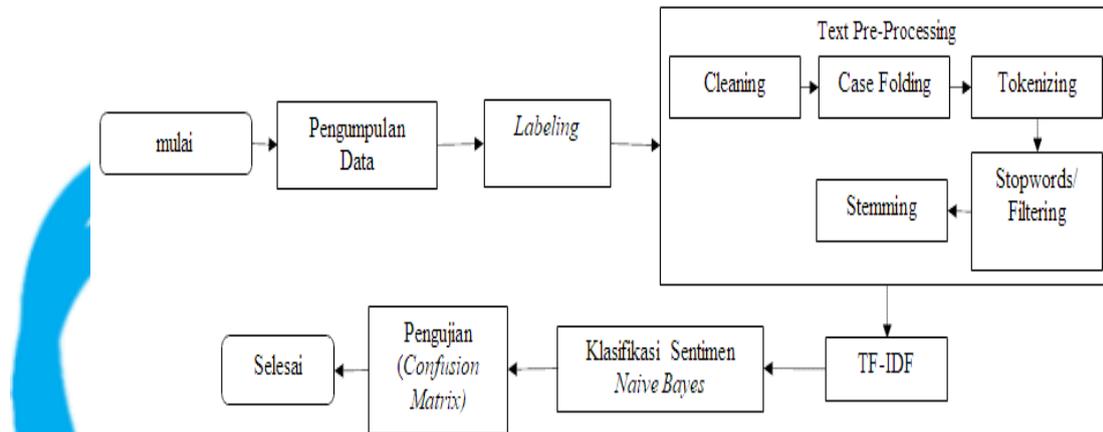
3.2 Peralatan Penelitian

Penelitian ini menggunakan berupa perangkat keras dan perangkat lunak sebagai berikut:

1. Perangkat Keras (*Hardware*)
 - a. Laptop (*Processor Inter Core i3-3200, 4GB RAM, 320GB HDD, Sistem Operasi Windows 10*)
 - b. *Mouse*
 - c. *Printer*
2. Perangkat Lunak (*Software*)
 - a. *Twitter* mencari data berupa cuitan atau *tweet*
 - b. *RapidMiner crawling data* dari *Twitter*
 - c. *Google Chorme* mencari referensi jurnal
 - d. *Microsoft Office* untuk membuat laporan

3.3 Prosedur Penelitian

Alur yang dilakukan dalam penelitian ini, di mulai dari proses pengumpulan data sampai pada tahap proses pengujian model algoritma. Dapat dilihat pada gambar alur penelitian dalam bentuk *Flowchart* ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Alur Penelitian

3.4 Pengumpulan Data

Penelitian ini mengambil data teks hasil cuitan *tweet* masyarakat tentang Vaksin Covid-19 di media sosial yaitu *Twitter*, dimana pengambilan data teks tersebut dilakukan dengan bantuan perangkat lunak yaitu *RapidMiner* hasil pengumpulan data ini disimpan ke dalam format *excel* (.xlsx). Proses pengumpulan data dengan *RapidMiner* ini mendapatkan 3069 *tweet*, kemudian pada proses klasifikasi data dibagi menjadi 2 yaitu data *training* sebanyak 80% dan data *testing* sebanyak 20%. Dengan lima (5) kata kunci “*Vaksin Sinovac*”, “*Vaksin Astrazeneca*”, “*Vaksin Moderna*”, “*Vaksin ke 3*”, dan “*Vaksin Booster*”.

3.5 Labelling

Hasil data yang didapat pada proses *crawling* akan diberikan label secara manual, menjadi 3 kelas sentimen yaitu sentimen negatif, sentimen netral dan sentimen positif.

3.6 Text Pre-processing

Dokumen teks ini masih mempunyai struktur atau kalimat yang tidak terstruktur dengan baik, diperlukan suatu proses yang dapat mengubah data tidak terstruktur menjadi data terstruktur dengan *text pre-processing*.

Text Pre-processing dilakukan bertujuan agar data yang digunakan bebas *noise* dan terstruktur agar dapat diproses lebih lanjut. Tahap *pre-processing* memiliki beberapa proses yaitu *cleaning*, *case folding*, *Stopword/Filtering*, *stemming*, dan *tokenizing*.

a. *Cleaning*

Cleaning, berfungsi untuk menghapus karakter atau atribut yang tidak diperlukan pada proses analisis berupa tanda baca, simbol-simbol, *hashtag*, *emoticon* dan *URL* dari situs web.

Tahap	Sebelum	sesudah
<i>Cleaning</i>	Dulu vaksin booster pake moderna, efeknya pusing, mual, panas dingin, sampe habisin paracetamol 5 tablet WKWKWK	Dulu vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet WKWKWK

Gambar 3. 2 Contoh Proses *Cleaning*

b. *Case Folding*

Case Folding, berfungsi untuk mengganti semua huruf *uppercase* atau huruf kapital menjadi *lowercase* atau huruf kecil.

Tahap	Sebelum	sesudah
<i>Case Folding</i>	Dulu vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet WKWKWK	dulu vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet wkwkwk

Gambar 3. 3 Contoh Proses *Case Folding*

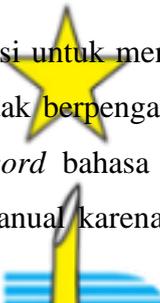
c. *Tokenizing*

Tokenizing, berfungsi untuk memisahkan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal.

Tahap	Sebelum	sesudah
<i>Tokenizing</i>	dulu vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet wkwkwk	['dulu', 'vaksin', 'booster', 'pake', 'moderna', 'efeknya', 'pusing', 'mual', 'panas', 'dingin', 'sampe', 'habisin', 'paracetamol', 'tablet', 'wkwkwk']

Gambar 3. 4 Contoh Proses *Tokenizing*d. *Stopwords Removing*

Stopwords Removing berfungsi untuk menyaring atau mengurangi kata yang tidak bermakna atau tidak berpengaruh dalam dokumen tersebut. Dengan menggunakan *stopword* bahasa Indonesia yang disusun dan menambahkan kata secara manual karena beberapa kata belum ada di *stopword*.



Tahap	Sebelum	sesudah
<i>Stopwords/ Filtering</i>	['dulu', 'vaksin', 'booster', 'pake', 'moderna', 'efeknya', 'pusing', 'mual', 'panas', 'dingin', 'sampe', 'habisin', 'paracetamol', 'tablet', 'wkwkwk']	vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet 'paracetamol', 'tablet', 'wkwkwk']

Gambar 3. 5 Contoh Proses *Stopword/Filtering*e. *Stemming*

Stemming, berfungsi untuk menghilangkan imbuhan (awalan dan akhiran) menjadi ke bentuk dasar pada kata dalam dokumen.

Tahap	Sebelum	sesudah
<i>Stemming</i>	vaksin booster pake moderna efeknya pusing mual panas dingin sampe habisin paracetamol tablet	vaksin booster pakai moderna efek pusing mual panas dingin sampai habis paracetamol tablet

Gambar 3. 6 Contoh Proses *Stemming***3.7 TF-IDF**

Setelah data bersih yang melalui tahapan *Text Preprocessing*, kemudian menghitung TF (*Trem Frequency*) dan IDF (*Inverse Document Frequency*) dengan menghitung bobot kata dan memberi nilai pada suatu kata untuk pada setiap dokumen.

3.8 Klasifikasi Sentimen *Naïve Bayes*

Metode *Naïve Bayes* metode yang digunakan untuk mengklasifikasikan data *tweet* untuk mendapatkan kategori sentimen analisis yang ditentukan. Klasifikasi sentimen menggunakan data yang diperoleh dari proses *preprocessing* hingga pembobotan kata oleh TF-IDF.

3.9 Pengujian

Pengujian dalam penelitian ini digunakan untuk mengevaluasi kinerja dari algoritma yang digunakan. Proses evaluasi dimulai dengan pembuatan *Confusion Matrix* untuk mengevaluasi keakuratan algoritma. Selanjutnya menguji kinerja algoritma menggunakan metode *Confusion Matrix*. *Confusion Matrix* ini sangat berguna untuk menganalisa kualitas sebuah klasifikasi.

Dalam *Confusion Matrix* dikenal beberapa istilah yang harus diketahui, yaitu: Tabel yang menggambarkan kinerja model atau algoritma tertentu. Setiap baris

TPos (<i>True Positive</i>)	=	kelas positif yang benar diklasifikasikan positif
FNeg (<i>False Negative</i>)	=	kelas negatif diklasifikasikan salah
FNet (<i>False Netral</i>)	=	kelas netral diklasifikasikan salah
TNet (<i>True Netral</i>)	=	kelas netral yang benar diklasifikasikan netral
TNeg (<i>True Negative</i>)	=	kelas negatif yang benar diklasifikasikan negatif
FP (<i>False Positive</i>)	=	kelas positif diklasifikasikan salah

Matriks mempresentasikan kelas data aktual, dan setiap kolom mewakili kelas data yang diprediksi (atau sebaliknya).

Tabel 3. 2 *Multiclass Confusion Matrix*

		S		
		Kelas Sebenarnya		
P		Negatif	Netral	Positif
		Kelas	Negatif	TNeg
Prediksi	Netral	FNet	TNet	FNet
	Positif	FPos	FPos	TPos

Rumus menghitung metode evaluasi dengan *Confusion Matrix* untuk menghitung *accuracy*, *precision*, *recall* dan *f1-score* seperti berikut:

Accuracy, perhitungan jumlah nilai positif diprediksi positif dan nilai negatif diprediksi negatif

$$accuracy = \frac{TNeg + TNet + TPos}{\text{Jumlah data testing}} \times 100\% \quad (5)$$

Precision, perhitungan ketepatan antara hasil prediksi dengan data sebenarnya

$$Precision = \frac{TPos}{TPos + FPos} \quad (6)$$

Recall, perhitungan tingkat keberhasilan sistem dalam memprediksi sebuah data

$$recall = \frac{TPos}{TPos + FN} \quad (7)$$

F1-score, perhitungan nilai perbandingan rata-rata *precision* dan *recall*

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

