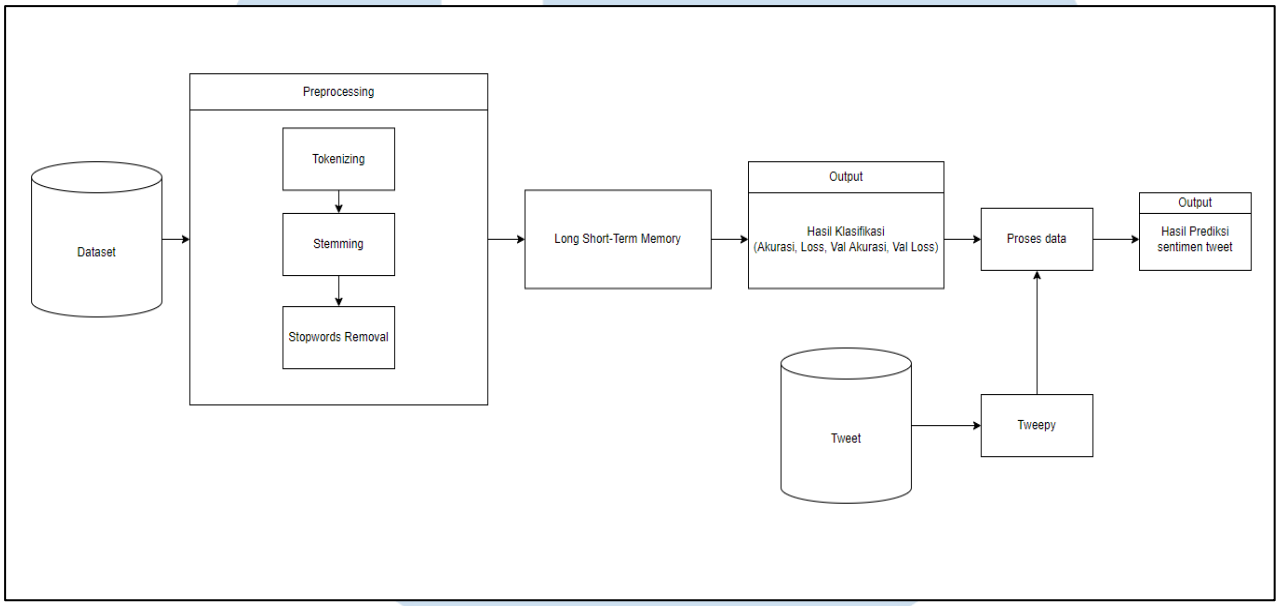
BAB 3

METODOLOGI PENELITIAN

3.1 Gambaran Jalan Sistem



Gambar 3.1 Diagram jalannya sistem

3.1.1 Data

Penelitian ini menggunakan *dataset* yang diambil dari jurnal.ugm.ac.id yang berisikan 10607 (sepuluh ribu enam ratus tujuh) contoh kalimat yang sudah dilengkapi dengan penanda sebagai -1 untuk kalimat dengan sentimen negatif, 0 untuk kalimat dengan sentimen netral, dan 1 untuk kalimat dengan sentimen positif. Lalu adapun data yang diteliti diambil dari *tweet* dengan topik PSBB tahap pertama dengan jangka waktu yang diambil dari tanggal 6 April 2020 hingga 14 April 2020. Semua *tweet* diambil langsung menggunakan python melalui library tweepy yang sebelumnya telah dikoneksikan dengan mencantumkan twitter API *key* dan twitter API *secret key* sebagai media autorisasi.

Data *tweet* dicari dan diambil melalui API twitter dengan ketentuan *tweet* merupakan data tanpa adanya tautan berupa gambar atau video, dan mengabaikan adanya sematan *hashtag* di dalam *tweet* tersebut. Sistem akan mengambil *tweet* sebanyak jumlah yang telah ditentukan user sebelumnya. Data yang didapat juga telah dipisahkan secara otomatis sehingga semua *tweet* yang masuk adalah *tweet* berbahasa Indonesia.

Tabel 3.1 Contoh Dataset *tweet* untuk training

Kategori	***Tweet***
Negatif	aku aja capek sama diriku sendiri apalagi kamu maaf ya
Netral	hi semuanya aku adain give away nih give away nya album bts yg love yourself answers yey cukup follow dan retweet
Positif	selamat hari jumat buat kamu yg selalu bikin rindu ini kumat
Negatif	aku tak faham betul lelaki tak suka perempuan pakai makeup
Netral	kalo aku bacanya sih emang kamu yang ngetik ga jelas
Positif	saat ditanya aku hanya diam dan mengangguk kutahu bahwa dia orang yang baik

3.1.2 Preprocessing

Tahap ini merupakan proses *indexing* pada *information retrieval* yang wajib dilalui oleh agar data yang masuk merupakan data-data yang cocok digunakan oleh pengguna. *Preprocessing* ini juga digunakan untuk mengklasifikasi data-data yang lebih spesifik. Adapun proses dari *preprocessing* dijabarkan sebagai berikut:

1. Menghitung panjang kalimat (*sentence length*)

   Di tahap ini, *tweet* dalam *dataset* akan dikumpulkan dan dihitung panjang dari setiap kalimat untuk mengukur apakah panjang dari sebuah *tweet* akan mempengaruhi hasil dari sentimen yang akan diprediksi oleh mesin. Perhitungan panjang kalimat disini akan dilakukan tiga kali yakni perhitungan dari *dataset* yang memiliki sentimen positif, *dataset* yang memiliki sentimen negatif, dan *dataset* yang memiliki sentimen netral.

2. *Tokenizing*

   Pada tahapan ini *tweet* dari dalam *dataset* akan dipecah setiap katanya ke dalam bentuk *token*, menghilangkan semua simbol, tanda baca, dan apapun yang tidak mewakili isi dokumen

   Adapun langkah dari *tokenizing* ialah sebagai berikut:

   a. Memproses teks *tweet* secara keseluruhan
   b. Mengambil *token* dari sebuah kalimat dengan spasi sebagai pemisah dari satu *token* dengan *token* lainnya serta melakukan *case-folding*
   c. Penghapusan simbol, tanda baca, serta hal lain yang tidak mewakili *tweet* tersebut

d. Menyimpan *token* ke dalam sebuah list sebagai satu *tweet*

Tabel 3.2 Contoh *Tokenizing Tweet* Negatif

Tweet	***Tokenizing***			
aku aja capek sama diriku sendiri apalagi kamu maaf ya	Aku	aja	capek	sama
	diriku	sendiri	apalagi	kamu
	maaf	ya		
aku tak faham betul lelaki tak suka perempuan pakai makeup	aku	tak	faham	betul
	lelaki	tak	suka	perempuan
	pakai	makeup		

Tabel 3.3 Contoh *Tokenizing Tweet* Netral

Tweet	***Tokenizing***			
hi semuanya aku adain give away nih give away nya album bts yg love yourself answers yey cukup follow dan retweet	hi	semuanya	aku	adain
	give	away	nih	give
	away	nya	album	bts
	yg	love	yourself	answers
	yey	cukup	follow	dan
	retweet			
kalo aku bacanya sih emang kamu yang ngetik ga jelas	kalo	aku	bacanya	sih
	emang	kamu	yang	ngetik
	ga	jelas		

Tabel 3.4 Contoh *Tokenizing Tweet* Positif

Tweet	***Tokenizing***			
selamat hari jumat buat kamu yang selalu bikin rindu ini kumat	selamat	hari	jumat	buat
	kamu	yang	selalu	bikin
	rindu	ini	kumat	
saat ditanya aku hanya diam dan mengangguk kutahu bahwa dia orang yang baik	saat	ditanya	aku	hanya
	diam	dan	mengangguk	kutahu
	bahwa	dia	orang	yang
	baik			

3. Stemming

Di tahap ini, dilakukan pembentukan kata yang memiliki imbuhan menjadi kata dasar dengan bantuan *library* Sastrawi. Sastrawi merupakan *free library* yang dapat digunakan untuk stemming kata-kata berbahasa Indonesia. Kata-kata yang dapat di-*stem* oleh Sastrawi berupa kata dengan imbuhan depan (prefiks), imbuhan belakang (sufiks), imbuhan sisipan di tengah kata dasar (infiks), imbuhan depan belakang (konfiks), serta pengulangan kata. Adapun hasil dari *stemming* menggunakan *library* sastrawi dijabarkan sebagai berikut:

Tabel 3.5 Contoh *Stemming Tweet* Negatif

Hasil *Tokenizing*	**Hasil *Stemming***
aku	aku
saja	saja
capek	capek

Hasil *Tokenizing*	**Hasil *Stemming***
sama	sama
diriku	diriku
sendiri	sendiri
apalagi	apalagi
kamu	kamu
maaf	maaf
ya	ya

Tabel 3.6 Contoh *Stemming Tweet* Netral

Hasil *Tokenizing*	**Hasil *Stemming***
kalo	kalo
aku	aku
bacanya	baca
sih	sih
emang	emang
kamu	kamu
yang	yang
ngetik	ketik
ga	ga
jelas	jelas

Tabel 3.7 Contoh *Stemming Tweet* Positif

Hasil *Tokenizing*	**Hasil *Stemming***
saat	saat
ditanya	tanya
aku	aku
hanya	hanya
diam	diam
dan	dan
mengangguk	angguk
kutahu	kutahu
bahwa	bahwa
dia	dia
orang	orang
yang	yang
baik	baik

4. *Stopword removal*

Pada tahap ini akan dilakukan penghilangan pada kata-kata yang tidak memiliki arti di bahasa Indonesia seperti kata depan, kata gabung dan lainnya dengan bantuan kamus *stopword* bahasa Indonesia. Berikut adalah contoh hasil *stopword removal*

Tabel 3.8 Contoh *Stopword Removal Tweet* Negatif

Hasil *Stemming*	**Hasil *Stopword removal***
aku	
aja	
capek	capek
sama	
diriku	diriku
sendiri	sendiri
apalagi	
kamu	
maaf	maaf
ya	

Tabel 3.9 Contoh *Stopword Removal Tweet* Netral

Hasil *Stemming*	**Hasil *Stopword removal***
kalo	
aku	
baca	baca
sih	
emang	
kamu	
yang	

Hasil *Stemming*	**Hasil *Stopword removal***
ketik	ketik
ga	
jelas	jelas

Tabel 3.10 Contoh *Stopword Removal Tweet* Positif

Hasil *Tokenizing*	**Hasil *Stemming***
saat	
tanya	tanya
aku	
hanya	
diam	diam
dan	
angguk	angguk
kutahu	kutahu
bahwa	
dia	
orang	orang
yang	
baik	baik

5. Visualisasi data menggunakan *Wordcloud*

Wordcloud adalah gambar yang menunjukkan daftar kata-kata yang digunakan dalam sebuah teks, umumnya semakin banyak kata yang digunakan semakin besar ukuran kata tersebut dalam gambar. Dalam sistem ini, *wordcloud* akan digunakan untuk memvisualisasikan kata-kata yang paling sering muncul di dalam sebuah kategori dataset. Tujuan dari pengadaan visualisasi data ini adalah untuk menunjukkan kata-kata yang muncul di sentimen negatif, positif, maupun netral.

3.1.3 Representasi Teks

Setelah mengalami preprocessing, *tweet* akan masuk ke dalam proses klasifikasi menggunakan LSTM. Langkah-langkah representasi teks adalah sebagai berikut:

1. *Bag of Words*

Bag of Words adalah metode paling sederhana dalam mengubah data teks menjadi vektor yang dapat dipahami oleh sebuah mesin. Metode ini akan menghitung frekuensi kemunculan dari kata pada seluruh *dataset*. Berikut adalah contoh *Bag of Words* yang telah diproses dari dataset, kata-kata yang ditampilkan disini merupakan 200 kata pertama yang berada di dalam *Bag of Words* dari *dataset tweet* yang telah disiapkan

['00' '1' '10' '100' '11' '12' '13' '14' '15' '17' '18' '2' '20' '2018'
'2019' '23' '24' '25k' '2k' '3' '30' '4' '5' '50' '50k' '59' '5k' '6' '7'
'8' '9' '90' 'a' 'aa' 'aamiin' 'ab' 'abai' 'abang' 'abg' 'abis' 'acappan'
'acara' 'acc' 'accident' 'ad' 'ada' 'adab' 'adain' 'ade' 'adek' 'adik'
'adu' 'aduh' 'aduk' 'ae' 'af' 'after' 'agama' 'agut' 'ah' 'ahhhhh' 'ai'

'air' 'aja' 'ajak' 'ajar' 'aji' 'ak' 'akak' 'aktif' 'akuekarangiapa'
'akun' 'al' 'ala' 'alam' 'alamat' 'alas' 'album' 'alhamdulillah' 'alim'
'all' 'allah' 'allahu' 'aloe' 'alone' 'also' 'am' 'ama' 'amal' 'ambik'
'ambil' 'ambyar' 'amek' 'amer' 'amik' 'amin' 'amp' 'ampai' 'ampe' 'ampun'
'an' 'ana' 'anai2' 'anak' 'and' 'anggap' 'angin' 'angkat' 'angkut'
'anjing' 'anon' 'apa' 'apaa' 'ape' 'api' 'app' 'appreciate' 'apps' 'arah'
'army' 'as' 'asa' 'asasi' 'asik' 'askmf' 'asli' 'assalam' 'astaga'
'asyik' 'at' 'atas' 'atleast' 'attack' 'aurat' 'auto' 'ava' 'awak' 'away'
'awek' 'ayah' 'ayam' 'ayat' 'ayo' 'b' 'bab' 'babi' 'babies' 'baby' 'baca'
'back' 'badan' 'badmood' 'bagi' 'bagiin' 'bagitahu' 'bagus' 'bahagia' 'bahasa'
'baik' 'baju' 'bakat' 'bal' 'balas' 'bales' 'bandareri'
'banding' 'bandung' 'bang' 'bangang' 'banget' 'bangga' 'bangun' 'bantu'
'bapa' 'barai' 'barang' 'barat' 'bareng' 'baring' 'barusan' 'bas' 'batas'
'bau' 'bawa' 'bawak' 'bawang' 'bayang' 'bayangebelum' 'bayar' 'bayi'
'bday' 'be' 'beb' 'bebas' 'beda' 'beg' 'bela' 'belah' 'belakang'
'belanja']

2. *Word Sequence*

Dari *dataset* yang telah dibersihkan, formatnya diubah menjadi tensor. Lapisan *embedding* akan menerima masukan berupa integer 2 dimensi. Dalam dataset ini terdapat 2705 kata yang telah disaring lalu diurutkan menggunakan *Bag of Words* diatas. Contoh representasi teks ditunjukan oleh tabel berikut

Tabel 3.11 Contoh Representasi teks

Sebelum *Tokenizing* & representasi teks	**Setelah *Tokenizing* & Representasi teks**
lagu bosan apa yang aku save ni huhuhuhuhuhuhuhuhuhuuuuuuuuuuuuuuu	[360 791 20 3 1 1465 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

3.2 Spesifikasi Sistem

Berikut adalah spesifikasi yang digunakan dalam pembuatan Analisis Sentimen Twitter Terhadap Tanggapan Mengenai Keputusan Masa PSBB dengan LSTM :

1. Perangkat lunak
   a. Sistem Operasi : Windows 10 64 bit
   b. Bahasa Pemrograman : Python 3.9.1
   c. Program : *Visual Studio Code*
2. Perangkat keras
   a. *Processor* : Intel Core i7-7700k CPU @2.80GHz
   b. RAM : 16 GB
   c. Hard disk : 1 TB