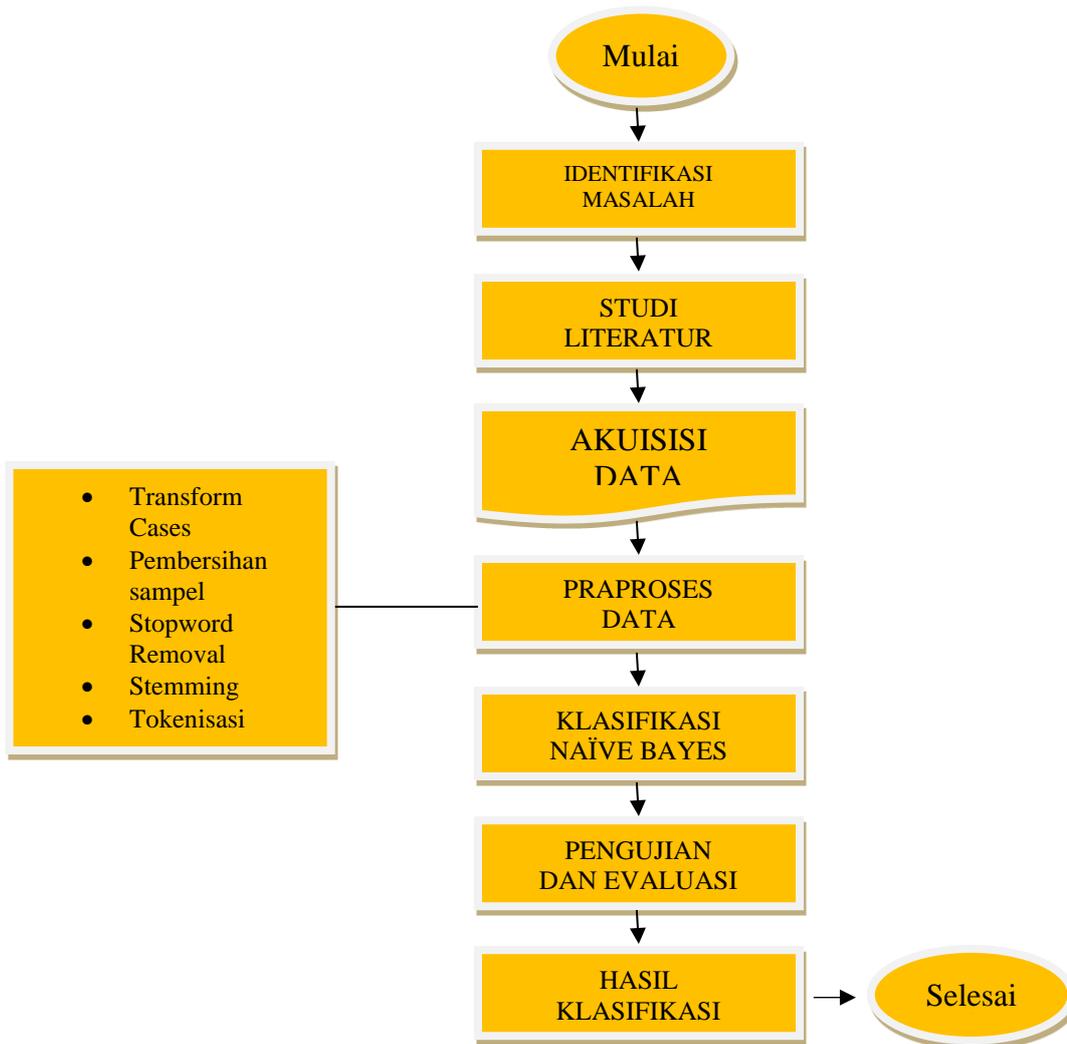


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Dalam mencapai tujuan dalam kegiatan penelitian yang akan dilakukan, penulis menyusun prosedur penelitian. Berikut merupakan tahapan penelitian yang dilakukan bisa dilihat pada Gambar.1.



Gambar 1. Tahapan Penelitian

Berikut merupakan metode yang digunakan dalam melakukan penelitian antara lain, identifikasi masalah, studi literatur, akuisisi data, pra proses data, proses, pemisahan sampel, pelatihan, pengujian dan dokumentasi.

3.1.1 Identifikasi masalah

Identifikasi masalah merupakan tahap untuk memperjelas masalah yang akan diangkat dalam penelitian ini yaitu permasalahan dalam mengklasifikasikan *tweet* yang mengandung opini tentang konflik yang terjadi antara Rusia dan Ukraina dengan metode klasifikasi *Naïve Bayes*.

3.1.2 Studi Literatur

Studi literatur dalam penelitian ini digunakan sebagai sumber pustaka dengan mengumpulkan buku jurnal terkait mengenai masalah mengenai analisis sentimen, text mining, metode klasifikasi, dan algoritma *Naïve Bayes* yang dibahas pada penelitian dan sebagainya. Pembahasan tersebut dicari dengan pengumpulan berbagai macam literatur, jurnal, *website* dan *e-book* yang berkaitan untuk penelitian. Setelah melakukan hal tersebut studi pustaka dijadikan acuan untuk memecahkan masalah pada penelitian ini.

3.1.3 Akuisisi Data

Akuisisi sampel dilakukan untuk mencari dan mendapatkan sampel yang dibutuhkan, sampel yang dimaksud dalam penelitian ini adalah *tweet* pengguna *twitter* yang berkaitan dengan konflik antara Rusia dan Ukraina yang akan dipakai untuk sampel latih dan sampel uji. Data *tweet* didapatkan dengan cara menggunakan API (*Application Programming Interface*) yang sudah disediakan oleh *twitter*.

3.1.4 Praproses Data

Pra proses sampel merupakan tahap awal yang dilakukan untuk mengolah sampel teks yang masih memiliki noise menjadi sampel yang siap untuk proses klasifikasi. Pada tahap ini juga bertujuan untuk mengambil inti dari sebuah teks supaya teks agar menjadi sampel yang siap dan tepat untuk diklasifikasikan.

3.1.5 Cleaning Data

Pembersihan sampel atau proses cleaning dilakukan untuk menghilangkan sampel yang tidak memiliki nilai atau informasi yang relevan. Dalam konteks ini, proses cleaning melibatkan penghapusan elemen-elemen tertentu seperti username, hashtag, URL, dan emotikon.

3.1.5.1 Stopword Removal

Stopword removal dilakukan untuk membuang kata-kata yang tidak memiliki kontribusi dalam proses klasifikasi (Fanissa, Fauzi, and Adinugroho 2018). Stopword sering muncul dalam jumlah besar dan tidak memiliki kontribusi pemaknaan. Contoh stopwords yang sering muncul adalah “yang”, “ke”, “dari”, dan lain-lain.

3.1.5.2 Stemming

Dalam proses pengklasifikasian kata-kata, ditemukan banyak kata yang memiliki imbuhan awalan maupun akhiran dari sebuah kata. Oleh karena itu, diperlukan adanya proses untuk menghilangkan imbuhan dari kata dasar agar makna yang dicari dalam proses klasifikasi lebih mudah diambil untuk selanjutnya diolah. Stemming merupakan proses untuk menentukan kata dasar dari kata yang berimbuhan (Zulfikar 2017). Sebagai contoh, proses stemming akan mendapatkan kata dasar dari “memarahi” yaitu “marah”.

3.1.5.3 Transform Cases

Tujuan dilakukannya *transform cases* yaitu untuk mengubah semua huruf pada susunan teks menjadi huruf kecil (*lowercase*) ataupun menjadi huruf besar (*uppercase*) (Utami, Lilyani Asri 2017). Dalam penelitian ini, semua huruf yang akan di klasifikasi diubah menjadi *lowercase* karena mayoritas *tweet* pada umumnya lebih banyak menggunakan *lowercase*. Contohnya adalah “Bekerja” menjadi “bekerja”.

3.1.5.4 Tokenisasi

Tokenisasi merupakan proses untuk memecah sebuah kalimat menjadi potongan karakter atau kata-kata sesuai dengan kebutuhan yang sering kita sebut token. Contohnya “sedang mengerjakan tugas”

menjadi “sedang” ‘mengerjakan’ ‘tugas’”.

3.1.6 Klasifikasi

Tahap selanjutnya dari penelitian ini yaitu pengklasifikasian sampel *tweet* yang telah melewati tahap pra proses sampel metode yang digunakan yaitu klasifikasi dan penelitian ini menggunakan *Naïve Bayes*. Algoritma tersebut akan mengklasifikasikan *tweet* yang mengandung opini terkait konflik antara Rusia dan Ukraina. Dalam pengerjaan menggunakan *Naïve Bayes* terdapat dua proses, yaitu sampel latih dan sampel uji. Langkah pertama ialah melakukan pelatihan sistem dengan sampel latih, langkah selanjutnya yaitu proses uji sistem dengan mengacu probabilitas sampel latih. Klasifikasi *Naïve Bayes* menggunakan library dari NLTK (*Natural Language Toolkit*) merupakan *library python* yang digunakan dalam proses *machine learning* teks.

3.1.7 Pengujian dan Evaluasi

Merupakan tahap untuk menguji hasil klasifikasi dengan menggunakan metode *confusion matrix* dengan sejumlah sampel yang diuji. Pada tahap ini menghitung nilai *accuracy*, *precision*, *recall*, dan *specificity*. “Evaluasi dilakukan untuk mengetahui nilai performansi dari sistem yang sudah dibuat berdasarkan hasil dari klasifikasi. Parameternya berdasarkan dari tabel *confusion matrix*. Tabel dibawah merupakan *confusion matrix* untuk klasifikasi dua *class*.” - (Novantirani, Sabariah, and Effendy 2015).

Tabel Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Sumber : <https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/>

Keterangan :

1. True Positive (TP), sampel yang diprediksi benar dan nyatanya benar
2. True Negative (TN), sampel yang diprediksi salah dan nyatanya salah.
3. False Positive (FP), sampel yang diprediksi benar dan nyatanya salah.
4. False Negative (FN), sampel yang diprediksi salah dan nyatanya benar.

Persamaan untuk menghitung nilai *accuracy*:

$$akurasi = \frac{TP+TN}{TP+FN+FP+TN}$$

Persamaan untuk menghitung nilai *precision*:

$$precision = \frac{TP}{TP+FP}$$

Persamaan untuk menghitung nilai *recall*:

$$recall = \frac{TP}{TP+FN}$$

Persamaan untuk menghitung nilai *specificity*:

$$specificity = \frac{TN}{TN+FP}$$

3.1.8 Hasil Klasifikasi

Pada tahap ini penulis akan mengumpulkan sampel yang diperlukan untuk selanjutnya dijadikan kesimpulan dalam penelitian ini. Setelah penulis mendapatkan kesimpulan, penulis bisa memberi saran untuk penelitian-penelitian berikutnya dan menjadi tahap akhir dalam penelitian ini.

3.2 Perangkat Penelitian

Dalam penelitian ini, alat-alat yang digunakan oleh penulis sebagai penunjang adalah:

a. Perangkat Keras

Perangkat keras yang digunakan adalah Laptop dengan spesifikasi sebagai berikut.

- i. Processor : AMD A10-9620P RADEON R5, 10 COMPUTE CORES 4C+6G, Graphics@2,5GHz
- ii. RAM : 8 GB
- iii. *Harddisk* : 1 TB

b. Perangkat Lunak

Perangkat lunak yang digunakan terdiri dari:

- i. Sistem Operasi : Microsoft Windows 10 Pro
- ii. Bahasa Pemrograman : Python
- iii. Aplikasi : Tweepy, Google Colab, Google Sheet

3.3 Waktu Kegiatan

Tabel 0.1. Waktu Kegiatan

Jadwal Pelaksanaan Penelitian	Bulan Ke-1				Bulan Ke-2				Bulan Ke-3				Bulan Ke-4				Bulan Ke-5				Bulan Ke-6			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Identifikasi Masalah																								
Studi Literatur																								
Akuisisi Data																								
Pra Proses Data																								
Proses Klasifikasi																								
Evaluasi																								
Dokumentasi																								