

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Subjek dan Objek Penelitian**

Objek penelitian merupakan suatu yang dijadikan fokus dalam melakukan suatu penelitian, yang dijadikan fokus dalam penelitian ini adalah mengenai analisis *Sea Games* 2021 Vietnam. Subjek dari penelitian ini adalah data tweet pengguna twitter yang menulis tweet terkait dengan *sea games* 2021.

#### **3.2. Alat dan Bahan**

Bahan penelitian yang digunakan dalam penelitian ini adalah data opini berbahasa Indonesia yang telah didapatkan dari sosial media twitter. Sedangkan alat yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

##### **3.2.1. Perangkat Keras ( *Hardware* )**

Perangkat keras yang dibutuhkan dan digunakan dalam pengembangan penelitian ini dengan sebuah laptop yang memiliki spesifikasi:

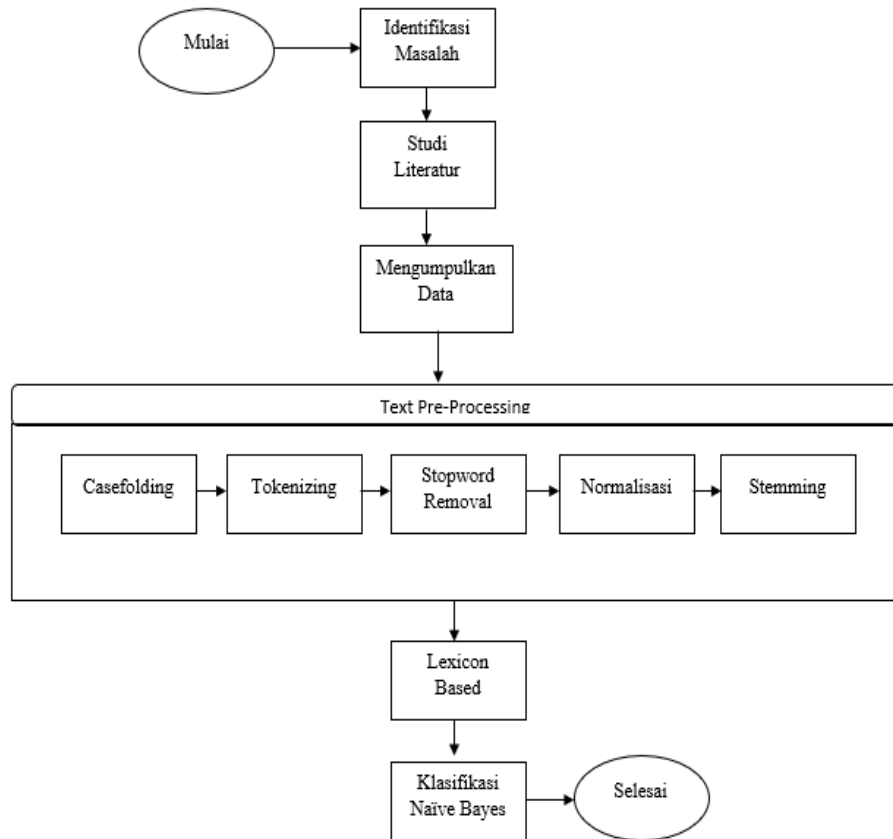
1. Laptop Asus dengan spesifikasi Intel Core I5.
2. Ram 4GB
3. Storage Hardisk 1 Tb
4. Graphics VGA NVIDIA GeForce 930MX

##### **3.2.2. Perangkat Lunak ( *Software* )**

Perangkat lunak yang dibutuhkan dan digunakan dalam pengembangan penelitian ini, sebagai berikut:

1. Sistem operasi windows 10
2. *Python 3.7 Command Prompt*
3. *Browser*
4. *Twitter*

### 3.3. Diagram Alir



**Gambar 3.1 Diagram Alir**

#### 3.3.1. Identifikasi Masalah

Langkah pertama yang dilakukan pertama kali yaitu mengidentifikasi dan merumuskan masalah tentang apa yang akan diteliti. Menentukan bidang, topik, masalah penelitian serta mengusulkan metode yang akan digunakan pada penelitian. Pada tahap ini mempelajari masalah yang masih terjadi di kehidupan sehari-hari, lalu tujuan, ruang lingkup serta metodologi penelitian.

#### 3.3.2. Studi Literatur

Langkah selanjutnya dengan melakukan studi literatur atau riset kepustakaan. Pada tahap ini, peneliti melakukan pengumpulan data – data yang berkaitan dengan topik

permasalahan yaitu tentang klasifikasi sentimen pengguna twitter, *Sea Games 2021*, tentang tweet di Twitter serta algoritma yang akan diusulkan yaitu naive bayes. Data – data untuk penelitian ini diperoleh dari jurnal, buku elektronik, dan media elektronik. Studi literatur memiliki tujuan untuk memperkuat permasalahan yang dibahas pada penelitian ini serta menjadi dasar untuk melakukan pengembangan selanjutnya.

### 3.3.3. Mengumpulkan Data

Langkah berikutnya adalah mengumpulkan data, data yang dikumpulkan berupa tweet pada Twitter tentang *sea games 2021*. Mengumpulkan data ini bertujuan agar mendapatkan informasi tentang *sea games 2021* ini untuk dapat di analisis dengan cara klasifikasi menggunakan metode *Naïve Bayes*. Pengumpulan data digunakan dengan melakukan *scrapping* data pada *tweet* yang ada di twitter.

**Tabel 3.1 Sampel Tweet Twitter**

No	Tweet
1.	Pemerintah bakal memberikan apresiasi berbentuk bonus kepada atlet peraih medali di ajang ASEAN Para Games 2022. Jumlah bonus yang akan diberikan diperkirakan bakal serupa dengan peraih medali di perhelatan SEA Games 2021. Sumber : Detik News #MGTRADIO #MGTNEWS #MGTPplay <a href="https://t.co/EeV6FKvEiV">https://t.co/EeV6FKvEiV</a>

2.	@BadmintonTalk Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di Vietnam ya? Atau Vietnam gak siap
3.	Siti Fadia yang memiliki nama lengkap Siti Fadia Silva Ramadhanti ini sudah banyak memenangkan kejuaraan Bulu Tangkis sejak tahun 2014 hingga kini.  Bersama Apriyani, penampilan Fadia langsung moncer saat meraih medali emas SEA Games 2021. <a href="https://t.co/HcX8y6Yxf4">https://t.co/HcX8y6Yxf4</a>
4.	Kesan saya, penghargaan berupa bonus kepada atlit Indonesia yang berprestasi (memperoleh medali) di era Jokowi sejak Asian Games 18 Jakarta, Olimpiade Tokyo, SEA Games Vietnam 2022 dan mungkin juga ASEAN Para Games 2022, jauh lebih baik dibanding dengan masa lalu.
5.	@tintinnieeee @UpdateBolabola Waktu sea games di Vietnam kemarin kalau ngga salah, Indonesia protes karena tempat latihan yang dikasih panpel Vietnam dinilai buruk. Sejak kejadian itu, Vietnam jadi sipaling kritis soal mengkritik kualitas lapangan kalau mereka main away.

Pada Tabel 3.1 diatas merupakan beberapa contoh data sampel *tweet* yang diambil dari media sosial *twitter*. Dimana data ini akan diolah menjadi analisis sentimen untuk melihat pesentasi dari pendapat masyarakat terhadap *Sea Games*.

#### 1.3.4. Text Pre-Processing

Langkah Langkah dalam melakukan pre-processing data adalah sebagai berikut :

##### 1. Case Folding

*Case Folding* adalah mengubah semua karakter huruf menjadi huruf kecil (*lowercase*). Contoh proses *Case Folding* dapat dilihat pada Tabel 3.2

**Tabel 3.2 Case Folding**

No	Sebelum Case Folding	Sesudah Case Folding
----	----------------------	----------------------

1	@BadmintonTalk Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di Vietnam ya? Atau Vietnam gak siap	bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di vietnam ya? atau vietnam gak siap
---	---	---

### 1. Tokenizing

*Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata seperti pada Tabel 3.3

**Tabel 3.3 Tokenizing**

No	Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
1	Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di Vietnam ya? Atau Vietnam gak siap	['Bukannya', 'tuan', 'rumah', 'asean', 'para', 'games', 'sama', 'dengan', 'sea', 'games', 'tahun', '2021', 'pelaksanaan', '2022', 'di', 'Vietnam', 'ya', 'Atau', 'Vietnam', 'gak', 'siap']

### 4. Stopword Removal

*Stopword Removal* yaitu proses penghapusan kata-kata yang terdapat pada *stoplist*. *Stoplist* itu sendiri berisi kosakata-kosakata yang bukan merupakan ciri dari suatu dokumen dan bertujuan agar menghilangkan kata yang tidak berguna. Contoh proses *Stopword Removal* dapat dilihat pada Tabel 3.4 dibawah ini.

**Tabel 3.4 Stopword Removal**

No	Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
1	@BadmintonTalk Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di Vietnam ya? Atau Vietnam gak siap	Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 di Vietnam ? Atau Vietnam gak siap

#### 5. Proses *Normalisasi*

Tahapan selanjutnya adalah proses *normalisasi*, sesuai dengan namanya pada proses ini bertujuan untuk menormalisasikan sebuah kata menjadi bentuk normalnya. Proses ini biasanya untuk menormalkan penulisan kata yang terdapat kesalahan penulisan (*typo*) ataupun kata yang disingkat. Seperti contoh apabila terdapat kata “*baikkk*” maka akan dinormalisasi menjadi kata “*baik*”. Contoh lain apabila terdapat kata yang disingkat seperti “*dln*” maka akan dinormalisasi menjadi “*dalam*”. Proses normalisasi ini juga dapat dilakukan menggunakan kamus normalisasi yang sudah tersedia.

#### 6. *Stemming*

*Stemming* adalah suatu proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). *Stemming* bertujuan untuk menghilangkan berbagai imbuhan seperti *prefiks*, *sufiks*, maupun *konfiks* yang ada pada setiap kata. Contoh proses *Stemming* dapat dilihat pada Tabel 3.5 dibawah ini.

**Tabel 3.5 *Stemming***

No	Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
----	-------------------------	-------------------------

1	Bukannya tuan rumah asean para games sama dengan sea games tahun 2021 (pelaksanaan 2022) di Vietnam ya? Atau Vietnam gak siap	Bukannya tuan rumah asean para games sama dengan sea games di Vietnam ya? Atau Vietnam gak siap
---	---	---

### 1.3.5. *Lexicon Based*

*Lexicon Based* digunakan untuk melakukan klasifikasi sentimen akan menggunakan data hasil proses *pre-processing*. Setelah data berhasil di *training* kemudian akan dilakukan pengujian menggunakan data test untuk menguji hasil ketepatan klasifikasi yang dilakukan.

### 1.3.6. **Klasifikasi menggunakan *Naïve Bayes***

Setelah melakukan *pre-processing* data, data tersebut masih bersifat mentah maka dibutuhkan label pada data tersebut yang bertujuan agar dapat melakukan klasifikasi. Setelah data mentah itu diolah menjadi data bersih, maka data tersebut bisa diolah sampai pelabelan, Proses pelabelan dimulai dengan melakukan pengecekan kata sentimen. Hasil akhir dari proses semua ini adalah akumulasi bobot atau polaritas kata yang terdapat di tiap data atau *tweet*.