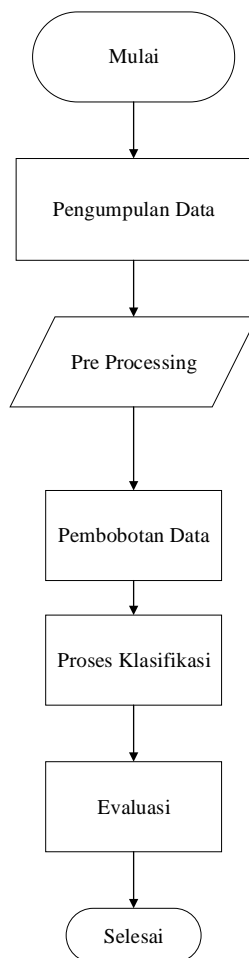


## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah dengan Penerapan metode Data Mining dengan Algoritma *Naïve Bayes* pada komentar Pengguna tiktok pada tahun 2022-2023 terhadap Pariwisata Lampung untuk di Analisa dan diklasifikasikan berdasarkan komentar yang Positif, Negatif dan Netral.

Berikut adalah Alur Penelitian menggunakan metode *Naïve Bayes* :



**Gambar 3.1 Alur Penelitian**

### 3.2 Tahapan Penelitian

Tahapan Penelitian yang akan dilakukan oleh peneliti, melalui beberapa tahapan penelitian. Tahapan penelitian adalah proses atau tata cara yang peneliti lakukan agar mendapatkan hasil yang diinginkan, adalah sebagai berikut:

#### 1. Studi Literatur

Tahapan pertama yang peneliti lakukan ialah Studi Literatur yaitu mencari, membahas dan memahami penelitian yang berkaitan dengan penelitian yang akan di bahas.

#### 2. *Crawling Data*

Tahapan selanjutnya yaitu *Crawling data*, data yang digunakan pada penelitian ini ialah data public yang diambil dari komentar pengguna TikTok terhadap Pariwisata Lampung menggunakan *TikTok Hastag Scrapper* yang selanjutnya akan di klasifikasikan menjadi kamus opini positif, negative dan netral.

#### 3. Preprocessing

Tahapan yang selanjutnya adalah *preprocessing*, dimana diharapkan dapat menghasilkan informasi yang dapat dianalisis, Informasi tersebut harus diproses terlebih dahulu untuk memastikan kualitas yang baik sebelum digunakan untuk analisis. Tujuan dari Prapemrosesan ini adalah untuk membersihkan dan menyiapkan data teks yang siap untuk di ekstraksi.

Tahapan Preprocessing yaitu sebagai berikut :

##### 1. *Cleaning*

Seringnya muncul symbol tanda baca dan bilangan angka pada komentar, membuat data menjadi tidak efektif dan tidak memiliki arti. Pada tahapan ini akan menghapus symbol, bilangan angka, dan tanda baca tersebut.

Tabel 3.1 Proses Cleaning Data

SEBELUM	SESUDAH
<p>Orang-orang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya</p> <p>#healingjourney #wisataindonesia #pariwisatalampung #lampung #wisatalampung #wonderfulindonesia #pantaimutun #dermagabomkaliandalampung #lembahbatuheritage #lengkunglangit2 #senjabeach #bendunganwaysekampung #pantaisebalang ##pantaikeduwarna #islamiccentertulangbawang barat #uluannughik</p>	<p>Orangorang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya</p> <p>healingjourney wisataindonesia pariwisatalampung lampung wisatalampung wonderfulindonesia pantaimutun dermagabomkaliandalampung lembahbatuheritage lengkunglangit senjabeach bendunganwaysekampung pantaisebalang pantaikeduwarna islamiccentertulangbawang barat uluannughik</p>
<p>Jalan lampung emang rusak tapi pantai lampung gak ada obat #pantailampung #wisatalampung #pariwisatalampung</p>	<p>Jalan lampung emang rusak tapi pantai lampung gak ada obat pantailampung wisatalampung pariwisatalampung</p>

## 2. Case Folding

Data yang digunakan ialah dari komentar di Tiktok, *Case Folding* adalah proses merubah huruf besar menjadi huruf kecil atau huruf standar. Proses ini mempermudah pencarian karena tidak semua dokumen teks memiliki kapitalisasi yang konsisten.

**Tabel 3.2 Proses Case Folding**

SEBELUM	SESUDAH
Orang-orang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya #healingjourney #wisataindonesia #pariwisatalampung #lampung #wisatalampung #wonderfulindonesia #pantaimutun #dermagabomkaliandalamp ung #lembahbatuheritage #lengkunglangit2 #senjabeach #bendunganwaysekampung #pantaisebalang ##pantaikeduwarna #islamiccentertulangbawang barat #uluannughik	orangorang lampung gk akan bingung buat cari tempat healing karna lampung tuh kaya akan tempat wisatanya healingjourney wisataindonesia pariwisatalampung lampung wisatalampung wonderfulindonesia pantaimutun dermagabomkaliandalamp ung lembahbatuheritage lengkunglangit senjabeach bendunganwaysekampung pantaisebalang pantaikeduwarna islamiccentertulangbawang barat uluannughik

Jalan lampung emang rusak tapi pantai lampung gak ada obat #pantailampung #wisatalampung #pariwisatalampung	jalan lampung emang rusak tapi pantai lampung gak ada obat pantailampung wisatalampung pariwisatalampung
---	--

### 3. Tokenization

Proses pengamanan data set dengan cara mengganti data-data pribadi sensitive dengan data informasi non-sensitif yang setara yang disebut dengan token dengan cara memisahkan kata satu persatu dalam bentuk koma.

**Tabel 3.3 Proses Tokenization**

SEBELUM	SESUDAH
Orang-orang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya #healingjourney #wisataindonesia #pariwisatalampung #lampung #wisatalampung #wonderfulindonesia #pantaimutun #dermagabomkaliandalamp ung #lembahbatuheritage #lengkunglangit2 #senjabeach #bendunganwaysekampung	['orangorang', 'lampung', 'gk', 'akan', 'bingung', 'buat', 'cari', 'tempat', 'healing', 'karna', 'lampung', 'tuh', 'kaya', 'akan', 'tempat', 'wisatanya', 'healingjourney', 'wisataindonesia', 'pariwisatalampung', 'lampung', 'wisatalampung', 'wonderfulindonesia', 'pantaimutun', 'dermagabomkaliandalampu ng', 'lembahbatuheritage', 'lengkunglangit', 'senjabeach',

#pantaisebalang ##pantaikeduwarna #islamiccentertulangbawan gbarat #uluannughik	'bendunganwaysekampung', 'pantaisebalang', 'pantaikeduwarna', 'islamiccentertulangbawang barat', 'uluannughik']
Jalan lampung emang rusak tapi pantai lampung gak ada obat #pantailampung #wisatalampung #pariwisatalampung	['jalan', 'lampung', 'emang', 'rusak', 'tapi', 'pantai', 'lampung', 'gak', 'ada', 'obat', 'pantailampung', 'wisatalampung', 'pariwisatalampung']

#### 4. Stopwords Removal

*Stopwords Removal* adalah proses menghilangkan kata-kata yang kurang penting atau menyesatkan dari kosakata yang tidak mempunyai arti.

**Tabel 3.4 Stopword Removal**

SEBELUM	SESUDAH
Orang-orang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya #healingjourney #wisataindonesia #pariwisatalampung #lampung #wisatalampung #wonderfulindonesia #pantaimutun #dermagabomkaliandalamp	['orangorang', 'lampung', 'gk', 'bingung', 'cari', 'healing', 'karna', 'lampung', 'tuh', 'kaya', 'wisatanya', 'healingjourney', 'wisataindonesia', 'pariwisatalampung', 'lampung', 'wisatalampung', 'wonderfulindonesia', 'pantaimutun', 'dermagabomkaliandalampu ng', 'lembahbatuheritage',

ung #lembahbatuheritage #lengkunglangit2 #senjabeach #bendunganwaysekampung #pantaisebalang ##pantaikeduwarna #islamiccentertulangbawan gbarat #uluannughik	'lengkunglangit', 'senjabeach', 'bendunganwaysekampung', 'pantaisebalang', 'pantaikeduwarna', 'islamiccentertulangbawang barat', 'uluannughik']
Jalan lampung emang rusak tapi pantai lampung gak ada obat #pantailampung #wisatalampung #pariwisatalampung	['jalan', 'lampung', 'emang', 'rusak', 'pantai', 'lampung', 'gak', 'obat', 'pantailampung', 'wisatalampung', 'pariwisatalampung']

### 5. Stemming

*Stemming* adalah mengurangi jumlah indikator yang berbeda dari informasi setelah ucapan yang diberi akhiran atau awalan ke bentuk dasarnya.

**Tabel 3.5 Proses Stemming**

SEBELUM	SESUDAH
Orang-orang lampung gk akan bingung buat cari tempat healing karna Lampung tuh kaya akan tempat wisatanya #healingjourney #wisataindonesia #pariwisatalampung	orangorang lampung gk bingung cari healing karna lampung tuh kaya wisata healingjourney wisataindonesia pariwisatalampung lampung wisatalampung wonderfulindonesia

<p>#lampung #wisatalampung  #wonderfulindonesia  #pantaimutun  #dermagabomkaliandalamp  ung #lembahbatuheritage  #lengkunglangit2  #senjabeach  #bendunganwaysekampung  #pantaisebalang  ##pantaikeduwarna  #islamiccentertulangbawang  barat #uluannughik</p>	<p>pantaimutun  dermagabomkaliandalamp  ung lembahbatuheritage  lengkunglangit senjabeach  bendunganwaysekampung  pantaisebalang  pantaikeduwarna  islamiccentertulangbawang  barat uluannughik</p>
<p>Jalan lampung emang rusak  tapi pantai lampung gak ada  obat #pantailampung  #wisatalampung  #pariwisatalampung</p>	<p>jalan lampung emang rusak  pantai lampung gak obat  pantai lampung  wisatalampung  pariwisatalampung</p>

#### 6. *Labeling Data*

Langkah selanjutnya adalah pelabelan pada kelas sentimen menggunakan kategori positif, negative dan netral. Pada penelitian ini, pelabelan data menggunakan 2 tahapan, sesuai dengan data yang digunakan pada penelitian ini sebanyak 1000 data maka pelabelan dilakukan dengan 70:30 yaitu 300 data akan dilabelkan menggunakan google collab dan 700 data dilabelkan menggunakan 3 ahli Bahasa Indonesia dengan masing masing ahli Bahasa melabelnya sebanyak 700 data untuk menghasilkan nilai akurasi yang lebih akurat.



### 7. *Splitting Dataset*

Data *splitting* merupakan proses membagi data menjadi data *training* (data latih) untuk mengajari sistem dan data *testing* (data uji) untuk menguji sistem. Algoritma Naïve Bayes yang digunakan dalam penelitian ini, termasuk ke dalam *supervised learning*. Sehingga tiap data yang digunakan untuk Training model adalah yang telah diberikan label

### 8. *Wordcloud*

Pada Tahap ini, tampilan bobot kata yang dominan muncul, divisualisasikan dalam bentuk wordcloud.