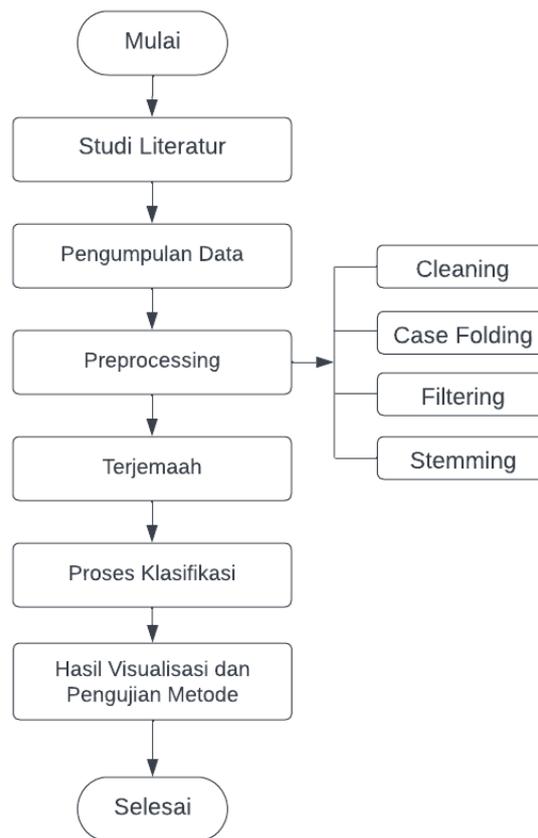


BAB III METODE PENELITIAN

3.1 Tahapan Penelitian

Pada Gambar 3.1 menunjukkan metode tahapan penelitian yang akan dilakukan :



Gambar 3.1 Tahapan Penelitian

3.1.1 Studi Literatur

Studi literatur dilakukan untuk mengumpulkan informasi dan referensi dari jurnal-jurnal atau karya ilmiah yang berkaitan dengan topik penelitian. Sumber-sumber tersebut meliputi konsep dan teori seperti Analisis Sentimen, Klasifikasi, dan Algoritma *Naive Bayes*.

3.1.2 Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah *web scraping*. *Web scraping* adalah gabungan metode yang digunakan untuk mendapatkan data dari situs web secara otomatis (Vargiu, 2012). Bahasa pemrograman yang digunakan untuk *web scraping* adalah Python. Dalam penelitian ini *tools* yang digunakan ialah *Google Colaboratory* dengan *library Google Play Scraper*.

3.1.3 Preprocessing

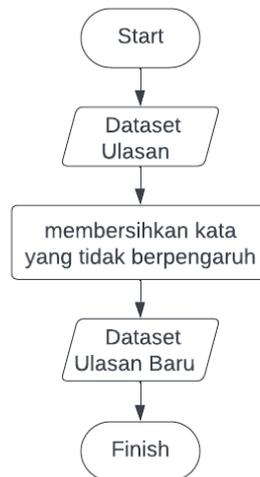
Setelah proses pengumpulan data dan disimpan dalam bentuk CSV maka langkah selanjutnya adalah *preprocessing data*, dikarenakan dataset tersebut belum terstruktur. Tugas utama *preprocessing data* adalah menghilangkan dan mengatasi *noise data* agar hasil perhitungan maksimal. Proses *preprocessing* terdiri dari beberapa tahapan yang dapat dilihat pada gambar 3.2 berikut.



Gambar 3.2 Diagram Alir *Preprocessing*

1. *Cleansing*

Pada tahap *cleansing* tanda baca dan karakter yang tidak diperlukan akan dihapus. Selain itu, karakter yang tidak relevan juga akan dihapus. Berikut diagram alir *cleansing* terdapat pada Gambar 3.3.



Gambar 3.3 Diagram Alir *Cleansing*

Sebagai gambaran dari proses *cleansing* berikut contoh ulasan yang dihasilkan seperti pada tabel 3.1

Tabel 3.1 Contoh Tahap *Cleansing*

Data Input	Data Output
Aplikasi nya sangat bagus sekali, Saya benar2 senang adanya Shopee ini karena kita di mudahkan dgn berbelanja online	Aplikasi nya sangat bagus sekali Saya benar senang adanya Shopee ini karena kita di mudahkan dgn berbelanja online

2. *Case Folding*

Pada tahap ini huruf kapital pada semua data ulasan akan diubah menjadi huruf kecil semua. Tujuannya untuk menghilangkan duplikasi data yang hanya berbeda pada hurufnya saja. Berikut diagram alir *case folding* terdapat pada Gambar 3.4.



Gambar 3.4 Diagram Alir *Case Folding*

Sebagai gambaran dari proses *Case Folding* berikut contoh ulasan yang dihasilkan seperti pada tabel 3.2

Tabel 3.2 Contoh Tahap *Case Folding*

Data Input	Data Output
Aplikasi nya sangat bagus sekali, Saya senang sekali adanya Shopee ini karena kita di mudahkan dengan berbelanja online dan banyak event event yang menarik	aplikasi nya sangat bagus sekali saya senang sekali adanya shopee ini karena kita di mudahkan dengan berbelanja online dan banyak event event yang menarik

3. *Tokenizing*

Tahap *tokenizing* ini adalah proses memisahkan kalimat menjadi setiap kata dan menyusunnya menjadi satu potongan. Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi . Hasil dari *tokenizing* ini adalah kata tunggal yang dimasukkan ke dalam database untuk proses pembobotan. Berikut diagram alir *tokenizing* terdapat pada Gambar 3.5.

Gambar 3.5 Diagram Alir *Tokenizing*

Sebagai gambaran dari proses *Tokenizing* berikut contoh ulasan yang dihasilkan seperti pada tabel 3.3

Tabel 3.3 Contoh Tahap *Tokenizing*

Data Input	Data Output
makin kesini subsidi ongkir jadi sedikit sekarang shopee makin mahal ongkirnya para pengguna jadi kecewa	'makin' kesini 'subsidi' 'ongkir' 'jadi' 'sedikit' 'sekarang' 'shopee' 'makin' 'mahal' 'ongkirnya' 'para' 'pengguna' 'jadi' 'kecewa'

4. *Filtering* atau *Stopwords Removal*

Dalam tahap *filtering* kata yang tidak memiliki arti akan dihilangkan agar terfokus pada kata-kata yang lebih bermakna. Sehingga proses selanjutnya yaitu klasifikasi akan lebih optimal karena jumlah kata yang diproses akan berkurang. Kata yang akan dihilangkan disimpan dalam *database* atau bisa *stopwordlist*. Jika dalam data ulasan ada kata yang sesuai dengan kata dalam

stopword maka kata tersebut akan dibuang dan diganti dengan spasi. Berikut diagram alir *filtering* pada Gambar 3.6.



Gambar 3.6 Diagram Alir *Filtering*

Sebagai gambaran dari proses *filtering* berikut contoh ulasan yang dihasilkan seperti pada tabel 3.4

Tabel 3.4 Contoh Tahap *Filtering*

Data Input	Data Output
makin kesini subsidi ongkir jadi sedikit sekarang shopee makin mahal ongkirnya para pengguna jadi kecewa	Subsidi ongkir sedikit ongkir mahal pengguna kecewa

5. *Stemming*

Pada tahap *stemming* kata berimbuhan akan diubah menjadi kata dasar.

Berikut diagram alir *stemming* pada Gambar 3.7.



Gambar 3.7 Diagram Alir *Stemming*

Sebagai gambaran dari proses *stemming* berikut contoh ulasan yang dihasilkan seperti pada tabel 3.5

Tabel 3.5 Contoh Tahap *stemming*

Data Input	Data Output
shopee sangat membantu <i>seller</i> untuk menjual barang	shopee bantu <i>seller</i> jual barang

3.1.4 Terjemaah

Setelah tahap *preprocessing* selesai, tahap berikutnya adalah menerjemahkan data ulasan bahasa Indonesia menjadi ulasan dalam bahasa Inggris. Ini dilakukan karena tahap klasifikasi berikutnya akan menggunakan *library TextBlob* dalam bahasa Inggris. Berikut diagram alir terjemaah pada Gambar 3.8.



Gambar 3.8 Diagram Alir terjemaah

Sebagai gambaran dari proses terjemaah berikut contoh ulasan yang dihasilkan seperti pada tabel 3.6

Tabel 3.6 Contoh Tahap Terjemaah

Data Input	Data Output
Aplikasi mudah digunakan. Dapatkan respons yang baik dan pengiriman yang sangat cepat	<i>App is easy to use. Do get good response and very fast delivery</i>

3.1.5 Proses Klasifikasi

Pada proses klasifikasi ini menggunakan Algoritma *Naive Bayes*. Dalam metode ini ada dua tahapan yaitu proses *training* dan proses *testing*. Proses *training* dilakukan terlebih dahulu sebelum proses *testing* dengan mempertimbangkan probabilitas dan dataset *training*.

Setelah *preprocessing* dan terjemaah selesai, langkah selanjutnya adalah pelabelan. Namun dalam penelitian ini dilakukan pelabelan otomatis dengan

Library TextBlob. API yang disediakan oleh *TextBlob* memungkinkan berbagai fungsi pemrosesan bahasa alami (NLP), termasuk analisis sentimen, klasifikasi, penerjemahan, dan ekstraksi frasa kata benda. Hasil objek *TextBlob* digunakan untuk memproses pembelajaran bahasa alami dan *library TextBlob* hanya dapat mengenali Bahasa Inggris.

Berikut perhitungan manual dari algoritma *Naïve Bayes Classification* dengan contoh 6 ulasan yang menjadi data *training* dan 3 ulasan data *testing*.

1. Proses *Training*

Sebanyak 6 ulasan yang menjadi data *training* sudah diklasifikasikan dan sudah dilakukan proses *preprocessing* secara manual seperti pada tabel 3.7 berikut:

Tabel 3.7 Contoh Kasus Data *Training*

Ulasan	Teks	Kelas
Ulasan1	<i>Good shopee service help free shipping</i>	P
Ulasan 2	<i>Okay thanks best discounts</i>	P
Ulasan 3	<i>Thanks help easy shopping</i>	P
Ulasan 4	<i>Sorry invoice not valid</i>	N
Ulasan 5	<i>Verry bad, expensive, slow</i>	N
Ulasan 6	<i>Application verry slow sorry</i>	N

Berdasarkan data tabel 3.7 dibuat sebuah model *probabilitas* dengan mengacu pada persamaan 2.7 dan 2.8 sebagai berikut:

$$P(a_thanks \mid v_positif) = (2 + 1)/(14 + 21) = 3/35$$

$$P(a_thanks \mid v_negatif) = (0 + 1)/(12 + 21) = 1/33$$

Berikut merupakan *probabilitas* setiap kata pada data *training* seperti pada tabel 3.8 :

Tabel 3.8 Perhitungan Probabilitas Data *Training*

Kelas	P(vj)	P(ai Vj)							
		good	shopee	service	help	free	shipping	okay	thanks
P	$\frac{1}{2}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{3}{35}$
N	$\frac{1}{2}$	$\frac{1}{33}$							

Kelas	P(vj)	P(ai Vj)							
		best	discount	easy	shopping	sorry	invoice	not	valid
P	$\frac{1}{2}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$
N	$\frac{1}{2}$	$\frac{1}{33}$	$\frac{1}{33}$	$\frac{1}{33}$	$\frac{1}{33}$	$\frac{3}{33}$	$\frac{2}{33}$	$\frac{2}{33}$	$\frac{2}{33}$

Kelas	P(vj)	P(ai Vj)				
		very	bad	expensive	slow	application
P	$\frac{1}{2}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$
N	$\frac{1}{2}$	$\frac{3}{33}$	$\frac{2}{33}$	$\frac{2}{33}$	$\frac{3}{33}$	$\frac{2}{33}$

Hasil perhitungan dari *probabilitas* tersebut diimplementasikan sebagai model probabilistik yang selanjutnya akan digunakan sebagai data acuan untuk menentukan data *testing*.

2. Proses *Testing*

Untuk perhitungan data *testing* menggunakan 3 sampel data ulasan, seperti pada tabel 3.9.

Tabel 3.9 Contoh Kasus Data Testing

Ulasan	Teks	Kelas
Ulasan7	<i>Thanks for free shipping</i>	?
Ulasan8	<i>Shopee easy used and help</i>	?
Ulasan9	<i>Product very expensive and bad</i>	?

Proses *testing* akan dihitung *probabilitasnya* dan dicari *probabilitas* tertinggi menggunakan persamaan 2.6 sebagai berikut:

$$\begin{aligned}
 &P(\text{Ulasan7}|V_{\text{positif}}) \\
 &= P(a_{\text{thanks}}|V_{\text{positif}}) \times P(a_{\text{for}}|V_{\text{positif}}) \times P(a_{\text{free}}|V_{\text{positif}}) \times \\
 &\quad P(a_{\text{shipping}}|V_{\text{positif}}) \times P(V_{\text{positif}}) \\
 &= \frac{3}{35} \times 1 \times \frac{2}{35} \times \frac{2}{35} \times \frac{1}{2} \\
 &= 0,0001399417
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Ulasan7}|V_{\text{negatif}}) \\
 &= P(a_{\text{thanks}}|V_{\text{negatif}}) \times P(a_{\text{for}}|V_{\text{negatif}}) \times P(a_{\text{free}}|V_{\text{negatif}}) \times \\
 &\quad P(a_{\text{shipping}}|V_{\text{negatif}}) \times P(V_{\text{negatif}}) \\
 &= \frac{1}{33} \times 1 \times \frac{1}{33} \times \frac{1}{33} \times \frac{1}{2} \\
 &= 0,0000139132
 \end{aligned}$$

$$\begin{aligned}
& P(\text{Ulasan8}|V_{positif}) \\
&= P(a_{shopee}|V_{positif}) \times P(a_{easy}|V_{positif}) \times P(a_{used}|V_{positif}) \times \\
&\quad P(a_{and}|V_{positif}) \times P(a_{help}|V_{positif}) \times P(V_{positif}) \\
&= \frac{2}{35} \times \frac{2}{35} \times 1 \times 1 \times \frac{1}{2} \\
&= 0,0016326531
\end{aligned}$$

$$\begin{aligned}
& P(\text{Ulasan8}|V_{negatif}) \\
&= P(a_{shopee}|V_{negatif}) \times P(a_{easy}|V_{negatif}) \times P(a_{used}|V_{negatif}) \times \\
&\quad P(a_{and}|V_{negatif}) \times P(a_{help}|V_{negatif}) \times P(V_{negatif}) \\
&= \frac{1}{33} \times \frac{1}{33} \times 1 \times 1 \times \frac{1}{2} \\
&= 0,00004591368
\end{aligned}$$

$$\begin{aligned}
& P(\text{Ulasan9}|V_{positif}) \\
&= P(a_{product}|V_{positif}) \times P(a_{very}|V_{positif}) \times P(a_{expensive}|V_{positif}) \times \\
&\quad P(a_{and}|V_{positif}) \times P(a_{bad}|V_{positif}) \times P(V_{positif}) \\
&= 1 \times \frac{1}{35} \times \frac{1}{35} \times 1 \times \frac{1}{35} \times \frac{1}{2} \\
&= 0,0000116618
\end{aligned}$$

$$\begin{aligned}
& P(\text{Ulasan9}|V_{negatif}) \\
&= P(a_{product}|V_{negatif}) \times P(a_{very}|V_{negatif}) \times P(a_{expensive}|V_{negatif}) \times \\
&\quad P(a_{and}|V_{negatif}) \times P(a_{bad}|V_{negatif}) \times P(V_{negatif}) \\
&= 1 \times \frac{33}{33} \times \frac{2}{33} \times 1 \times \frac{2}{33} \times \frac{1}{2} \\
&= 0,0001669588
\end{aligned}$$

Setelah menghitung *probabilitas* dari setiap data *testing*, diperoleh hasil pada tabel 3.10 sebagai berikut:

Tabel 3.10 Nilai Probabilitas Data Testing

Ulasan	Probabilitas	
	Positif	Negatif
Ulasan7	0,0001399417	0,0000139132
Ulasan8	0,0016326531	0,00004591368
Ulasan9	0,0000116618	0,0001669588

Pada tabel 3.10 dapat disimpulkan hasil dari data *testing* pertama yaitu Ulasan7 didapatkan nilai *probabilitas* positif lebih besar dibandingkan nilai *probabilitas* negatif sehingga dapat disimpulkan bahwa Ulasan7 termasuk ke dalam kategori sentimen positif. Pada Ulasan8 didapatkan nilai *probabilitas* positif lebih tinggi dari nilai *probabilitas* negatif sehingga Ulasan8 juga termasuk ke kategori sentimen positif. Sedangkan Ulasan9 nilai *probabilitas* negatif nya lebih unggul jika dibandingkan dengan nilai *probabilitas* positif sehingga Ulasan9 dikategorikan sentimen negatif.

3.1.6 Hasil Visualisasi dan Pengujian Metode

Hasil dari tahap visualisasi dalam penelitian ini digambarkan dengan *wordcloud*. Tampilan *wordcloud* akan menampilkan kata yang sering muncul dalam masing-masing sentimen.

Pengujian metode dilakukan untuk mengevaluasi kinerja model. Pengujian ini dilakukan dengan menilai tingkat akurasi menggunakan confusion matrix untuk setiap model. Setelah data *testing* diujikan dengan data *training*, akan dihasilkan daftar kelas-kelas dari data *testing*, dikenal sebagai data prediksi kelas. Kemudian prediksi kelas dibandingkan dengan kelas yang sebenarnya dari data *testing*. Sehingga dapat dilihat dan dihitung nilai *accuration*, *precision*, *recall*, dan *f1-score*.

3.1 Objek Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data primer. Menurut Sekaran (2011), data primer berarti informasi yang telah dikumpulkan langsung oleh peneliti dan terkait dengan variabel minat untuk tujuan studi tertentu. Data untuk penelitian ini diperoleh melalui teknik *web scraping* dari *Google Play Store* shopee menggunakan *Tools Google Colab* dan *library Google Play Scraper*. Data yang diperoleh berupa komentar atau ulasan pengguna aplikasi *Shopee* yang diambil sejak tanggal 2 November 2022 sampai tanggal 31 Januari 2023 dengan total 4902 ulasan.

3.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah *database website Google Play*, yaitu semua data ulasan *Shopee* pada bulan November 2022 hingga Januari 2023. Sedangkan untuk sampel yang digunakan adalah ulasan *Shopee* dari sejak pertengahan *upgrade* aplikasi *Shopee* yaitu tanggal 2 November 2022 hingga terakhir *upgrade* pada akhir Januari 2023.