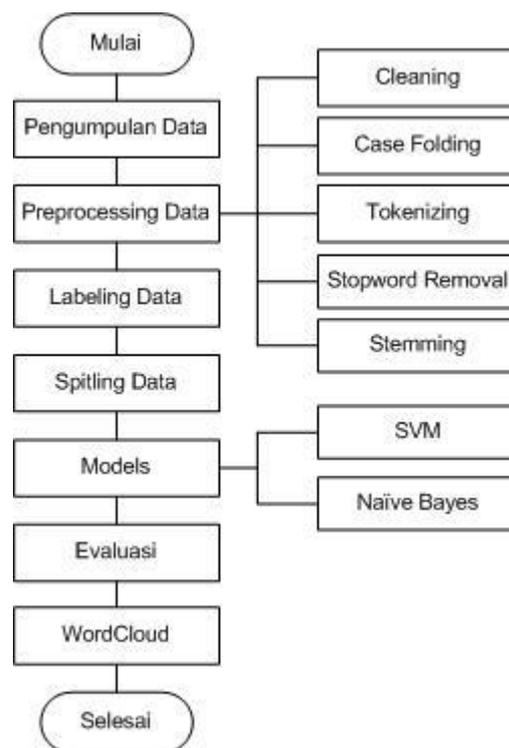


## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Dalam penelitian ini, metode yang digunakan meliputi *Support Vector Machines (SVM)* dan *Naïve Bayes*. Tahapan-tahapan dalam melakukan penelitian ini disajikan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

### 1.2 Pengumpulan Data

Pada tahap ini, proses pengumpulan data dilakukan dengan menggunakan data skunder yang diperoleh dari ulasan pengguna aplikasi Seabank. Dataset yang berhasil dikumpulkan terdiri dari 3.804 ulasan yang diperoleh dari ulasan aplikasi Seabank di Google Play Store dengan kategori komentar yang terbaru, rentang waktu dari 01 Januari 2022 hingga 20 Desember 2023. Pengumpulan data dilakukan melalui teknik web scraping dengan

menggunakan metode perhitungan menggunakan Google Colab dan diprogram dengan bahasa Python. berikut hasil scrapping data ulasan pengguna aplikasi seabank dalam format csv dapat dilihat pada gambar 3.2

reviewId	userName	userImage	content	score	thumbsUp	reviewCreat	replyContent	repliedAt
106568ba-bf2d-4e45-911b-9c	Fikri Fahmi Udin	https://play-lh.google	Jangan sampai naruh duit banyak, bahaya	1	2645	2.60.0	10/12/2023 05:22	Hai Sobat SeaBank, m
e1f2b9e6-39b1-4d07-bce8-f5	Dian Mutamimah	https://play-lh.google	Aplikasi sampah, saya telpon beberapa l	1	1500	2.59.0	28/11/2023 12:14	Hai Sobat SeaBank, m
29a79e6-f16-4293-9fb9-dff	dita heydier	https://play-lh.google	Aplikasi gak ada GUNA. Kalo terjadi kenc	1	385	2.59.0	01/12/2023 15:31	Hai Sobat SeaBank, m
a834104f-351d-4960-ab78-9a	desi anandaas	https://play-lh.google	bad service handle satu kendala hampir	1	255	2.58.0	10/11/2023 06:31	Sebelumnya mohon
f32708e-26af-445f-b705-b3	dunia jeje	https://play-lh.google	Aplikasi yang tidak bisa kasih solusi, Kon	1	530	2.60.0	08/12/2023 05:48	Hai Sobat SeaBank, m
7d015df5-0817-4b4f-ab7d-90	TAZALCY DUARIBUE	https://play-lh.google	Akun Seabank saya di blokir sedangkan :	1	326	2.59.0	04/12/2023 03:12	Hai Sobat SeaBank, m
bf72532a-928d-4bcf-bb5c-88	Azmi Maliki	https://play-lh.google	tolong jika memang ada gangguan dalam	2	156	2.59.0	20/11/2023 06:30	Hi Sobat SeaBank, m
686b020-7716-4e90-9523-c4	vir go	https://play-lh.google	apk ter bagus menurut saya karena buat t	5	93	2.60.0	12/12/2023 07:43	Halo Sobat SeaBank,
20c6c87-946c-4534-a0c1-88c	Primarini Riati	https://play-lh.google	Setiap menghubungi CS baru berapa me	1	136	2.60.0	11/12/2023 08:09	Hai Sobat SeaBank, m
7bb4ca3a-eea4-48a4-8de9-1c	Marulli Ruli	https://play-lh.google	Hari ini saya sangat kecewa lagi di mana	1	75	2.58.0	09/11/2023 04:10	Hai Sobat SeaBank, m
b89c90e-1b44-4c59-b005-2e	Hadelaya Channel	https://play-lh.google	Tadinya mau menjadikan seabank sebag	3	37	2.60.0	18/11/2023 02:59	Hi Sobat SeaBank, m
e8acd53a-6210-4d29-928c-14	Hafizh Dannur Hafshz	https://play-lh.google	Overall UI aplikasi sudah bagus dan mud	3	100	2.58.0	12/11/2023 02:01	Hi Sobat SeaBank, m
2b9e2d17-7441-447f-99d7-3e	Alya Vinoata	https://play-lh.google	sy sudah isi saldo seabank, mau chekut	3	0	2.59.0	19/12/2023 08:08	Hai Sobat SeaBank, m
f507074-7631-4965-a99f-a45	Alya Ovia Sinaga	https://play-lh.google	first time to use Seabank as saving, sem	5	2	2.60.0	18/12/2023 00:55	
959890ef-d801-4776-a78c-5c	Rahmat	https://play-lh.google	Aplikasi SIAL memang, tidak pernah sala	1	5	2.60.0	18/12/2023 05:11	Hi Sobat SeaBank, m
7f253f32-41ff-4389-9faa-513	Della Amanda	https://play-lh.google	Disuruh hubungkan rekening sea bank d	1	3	2.60.0	17/12/2023 19:24	Hai Sobat SeaBank, m
d1e4c3ce-7a78-4d8d-9511-b3	Semangka L	https://play-lh.google	Ga percaya lagi sama Seabank. Bisa-bisa	1	39	2.58.0	16/11/2023 12:27	Hai Sobat SeaBank, m
392aad3c-5b62-49ad-bc41-61	Miftha B	https://play-lh.google	Mau login disuruh verifikasi wajah lagi, st	1	50	2.59.0	26/11/2023 09:15	Hai Sobat SeaBank, m
9844d43e-e266-4607-8074-57	Sevi Haryanti	https://play-lh.google	Sejauh ini bagus, transaksi dll tanpa ken	5	5	2.61.0	20/12/2023 12:20	Hai Sobat SeaBank, m
6e250d3e-clf0-4e38-9c07-b5	ubur ubur	https://play-lh.google	Awal2 launching, aman bagus2 aja app n	1	66	2.60.0	13/12/2023 21:15	Hi Sobat SeaBank, m

Gambar 3.2 Hasil *web scrapping* data dalam format CSV

### 1.3 Preprocessing Data

Tahap berikutnya adalah *preprocessing*, yang bertujuan untuk menghasilkan informasi yang dapat dianalisis dengan baik. Data harus diproses terlebih dahulu untuk memastikan kualitasnya sebelum digunakan untuk analisis. Tujuan dari *preprocessing* adalah membersihkan dan menyiapkan data teks agar siap untuk diekstraksi fitur. Tahap data *preprocessing* ini terdiri dari proses *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal* dan *Stemming* [19].

#### 1. *Cleaning*

Adanya simbol, tanda baca, dan angka dalam komentar sering membuat data kurang efektif dan bermakna. Pada tahap ini, simbol, angka, dan tanda baca tersebut akan dihapus. Contoh penerapan tahap pembersihan ini dapat dilihat pada Tabel 3.1.

Tabel 3.1 Contoh *Cleaning*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 🙄	Aplikasi gak ada GUNA Kalo terjadi kendala susah komplainnya Udah habis pulsa saya ribu lebih gak menemukan solusi CS ngomong lambat banget bertele tele masalah tidak terselesaikan Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya Tidak recommended

## 2. Case Folding

Data yang digunakan berasal dari komentar di Google Play. Case Folding adalah proses mengubah huruf besar menjadi huruf kecil atau standar, seperti yang ditunjukkan dalam Tabel 3.2.

Tabel 3.2 Contoh *Case Folding*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 🙄	aplikasi gak ada guna kalo terjadi kendala susah komplainnya udah habis pulsa saya ribu lebih gak menemukan solusi cs ngomong lambat banget bertele tele masalah tidak terselesaikan buang buang waktu dan uang aja buat beli pulsa ngehubungin cs nya tidak recommended

### 3. *Tokenizing*

*Tokenizing* adalah tahapan yang bertujuan untuk membagi teks dari ulasan menjadi potongan-potongan kata, yang disebut sebagai token. Tujuan dari *tokenizing* adalah agar data dapat diproses lebih lanjut pada tahap selanjutnya, yaitu penghapusan *stopwords*. Contoh penerapan *tokenizing* dapat dilihat pada Tabel 3.3.

Tabel 3.3 Contoh *tokenizing*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 🙄	[aplikasi,gak,ada,guna,kalo,terjadi ,kendala,susah,komplainnya,udah, habis,pulsa,saya,ribu,lebih,gak,menemukan,solusi,cs,ngomong,lambat,banget,bertele,tele,masalah,tidak,terselesaikan,buang,buang,waktu,dan,uang,aja,buat,beli,pulsa,ngehubungin,cs,nya,tidak,recommended]

### 4. *Stopwords Removal*

*Stopwords Removal* adalah proses menghilangkan kata-kata yang kurang penting atau bisa menyesatkan dari kosakata, karena kata-kata tersebut tidak memiliki arti yang signifikan. Contoh penerapan penghapusan *stopwords* dapat dilihat pada Tabel 3.4.

Tabel 3.4 Contoh *Stopwords Removal*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah	[aplikasi,gak,kalo,kendala,susah,komplainnya,udah,habis,pulsa,ribu,gak,menemukan,solusi,cs,ngomong,lambat,banget,bertele,tele,terselesaikan,buang,buang,uang,aja,beli]

tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 🙄	,pulsa,ngehubungin,cs,nya,recomended]
---	---------------------------------------

### 5. *Stemming*

*Stemming* adalah mengurangi jumlah indikator yang berbeda dari informasi setelah ucapan yang diberi akhiran atau awalan ke bentuk dasarnya. Pada tahap ini data reviews yang telah bersih proses kembali dengan melakukan stemming untuk mengganti kata berimbuhan menjadi kata dasar untuk memudahkan proses penghilangan kata tidak bermakna. dapat dilihat Pada tabel 3.5.

Tabel 3.5 Contoh *Stemming*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 🙄	aplikasi gak kalo kendala susah komplain udah habis pulsa ribu gak temu solusi cs ngomong lambat banget tele tele selesai buang buang uang aja beli pulsa ngehubungin cs nya recomended

## 1.4 Labeling Data

Tahap berikutnya pelabelan pada kelas sentimen dengan menggunakan kategori positif, netral, dan negatif. Pada pelabelan ini menggunakan 2 tahapan yaitu google colab dan ahli bahasa. Data yang diperoleh dari hasil web scraping URL aplikasi Seabank di Google Play Store.

## 1.5 Splitting Dataset

Data splitting merupakan tahapan di mana data dibagi menjadi data latih (training data) dan data uji (testing data). Jumlah data latih menunjukkan berapa banyak sampel data yang digunakan untuk melatih model, sementara jumlah data uji menunjukkan berapa banyak sampel yang digunakan untuk menguji kinerja model. Setiap data yang digunakan untuk melatih model telah diberi label, dan dalam penelitian ini, dilakukan pembagian dengan rasio 80:20, yang berarti 80% data digunakan untuk training dan 20% digunakan untuk testing. [21]

## 1.6 Models

Pada tahap pemodelan dalam penelitian ini, digunakan dua model klasifikasi, yaitu metode *Naïve Bayes* dan *Support Vector Machine*.

### 1. *Naïve Bayes*

Metode *Naïve Bayes* merupakan salah satu algoritma yang digunakan dalam klasifikasi. Algoritma ini menggunakan probabilitas dan statistik yang dikembangkan oleh ilmuwan Inggris Thomas Bayes, yang bertujuan untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal sebagai Teorema Bayes. Dalam penelitian ini, digunakan tiga jenis model *Naïve Bayes*, yaitu model *Gaussian Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Bernoulli Naïve Bayes* [19]

### 2. *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) adalah algoritma klasifikasi yang cepat dan sering digunakan karena menghasilkan hasil yang baik. Dalam SVM, terdapat garis pemisah yang disebut hyperplane yang digunakan untuk memisahkan sentimen positif, netral, dan negatif [19].

## 1.7 Evaluasi

Pengujian pada penelitian ini mencakup evaluasi hasil dari algoritma yang digunakan. Metode evaluasi yang digunakan adalah *Confusion matrix*, yang sangat berguna untuk menganalisis kualitas klasifikasi. *Confusion matrix* merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Tabel pada *confusion matrix* menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah diklasifikasikan. Pengujian akan menghitung nilai *akurasi*, *recall*, *presisi*, dan *f1-score*, yang akan ditampilkan dalam bentuk persentase. Perhitungan matriks konfusi digunakan untuk menghitung akurasi pengujian, dan dilakukan sebagai berikut. [19].

1. Akurasi adalah proporsi keseluruhan prediksi yang benar.

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

2. Nilai precision atau presisi dilakukan untuk mengukur seberapa akurat pengujian data dalam memprediksi sentimen. Presisi didapatkan dari rasio jumlah prediksi benar dibagi dengan jumlah total prediksi.

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

3. Nilai racall atau sensitivitas dilakukan untuk mengukur seberapa baik pengujian data dapat mengenali semua instance data. Recall didapatkan dari rasio jumlah prediksi benar dibagi jumlah total nilai aktual.

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

4. Nilai F1-score dilakukan untuk memberikan gambaran yang seimbang tentang performa model dalam memprediksi dan mengenali data sentimen. Nilai F1-score didapatkan dari hasil perhitungan label prediksi (precision) dan label aktual (recall).

$$F1 - Score = 2 \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

TP : Nilai True Positive

FP : Nilai False Positive

FN : Nilai False Negative

TN : Nilai True Negative

## 1.8 Word Cloud

Pada tahap ini, dilakukan visualisasi kata-kata yang sering muncul dalam teks dengan menggunakan ukuran font yang lebih besar.