

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Gambaran Umum Objek Penelitian**

Penelitian ini secara khusus menyoroti analisis sentimen tanggapan pengguna aplikasi Halo BCA yang diambil dari ulasan pada platform Google Play Store. Ulasan yang dikumpulkan mencakup tanggapan pengguna terhadap performa aplikasi, kualitas layanan, serta pengalaman penggunaan aplikasi tersebut. Rentang waktu penelitian ini melibatkan data ulasan yang dikumpulkan dalam periode tertentu guna memastikan keakuratan dan relevansi data yang dianalisis.

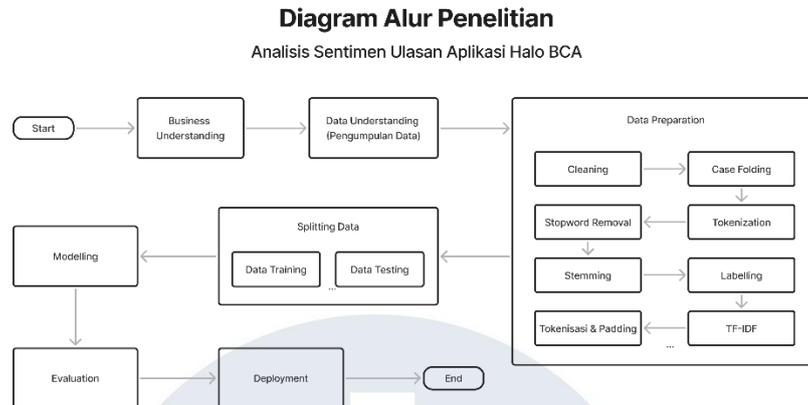
Pendekatan yang digunakan melibatkan perbandingan kinerja tiga algoritma klasifikasi sentimen, yaitu Support Vector Machine (SVM), Naïve Bayes, dan Long Short-Term Memory (LSTM). Data yang diperoleh akan melalui tahap preprocessing untuk meningkatkan kualitas input sebelum dilakukan pemodelan dan evaluasi performa metode yang digunakan.

Implementasi analisis ini dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa pustaka seperti Scikit-learn untuk membangun model klasifikasi, NLTK untuk pemrosesan bahasa alami, serta Jupyter Notebook sebagai lingkungan pengembangan untuk menjalankan dan menganalisis hasil penelitian secara interaktif.

#### **3.2 Metode Penelitian**

##### **3.2.1 Alur Penelitian**

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk memastikan keakuratan dan validitas hasil yang diperoleh.



Gambar 3. 1 Diagram Alur Penelitian

Gambar 3.1 merupakan diagram alur penelitian ini yang diadopsi dan disesuaikan dari diagram alur penelitian yang telah digunakan dalam penelitian terdahulu agar relevan dengan tujuan dan metodologi yang diterapkan [31]. Alur penelitian meliputi:

1. *Business Understanding*

Penelitian ini menganalisis sentimen ulasan pengguna aplikasi Halo BCA menggunakan metode LSTM, Naïve Bayes, dan SVM untuk membantu peningkatan layanan digital perbankan.

2. *Data Understanding (Pengumpulan Data)*

*Dataset* yang digunakan dalam penelitian ini bersumber dari ulasan pengguna aplikasi Halo BCA yang dikumpulkan melalui teknik *web scraping* dari Google Play Store

3. *Data Preparation*

Data ulasan yang diperoleh akan melalui tahap *cleaning*, *case folding*, *tokenization*, *stopword removal*, *stemming*, *labelling*, pembobotan TF-IDF, Tokenisasi dan *Padding*.

4. *Splitting Data*

Data ulasan aplikasi Halo BCA dibagi menjadi data latih (*training*) dan data uji (*testing*) untuk melatih model klasifikasi dan mengukur kinerjanya secara objektif dalam analisis sentimen.

### 5. *Modelling*

Data yang telah diproses digunakan untuk pelatihan model SVM, Naïve Bayes dan LSTM dalam Python menggunakan pustaka seperti Scikit-learn dan NLTK.

### 6. *Evaluation*

Evaluasi terhadap model dilakukan dengan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score untuk membandingkan efektivitas masing-masing algoritma.

### 7. *Deployment*

Tahapan ini bertujuan agar model dapat digunakan untuk memantau dan mengevaluasi sentimen pengguna secara otomatis sehingga mendukung pengambilan keputusan dan peningkatan layanan berbasis data.

## 3.2.2 Metode Data Mining

Penelitian ini membandingkan tiga metode *data mining*, yaitu CRISP-DM, SEMMA, dan KDD, untuk melihat perbedaan tahapan dan pendekatannya dalam proses penggalian data. Berikut adalah tabel perbandingan ketiga metode tersebut:

Tabel 3. 1 Perbandingan Metode Data Mining [30]

Metode	CRISP-DM	SEMMA	KDD
Kelebihan	1. Struktur proses yang jelas 2. Fleksibilitas 3. Fokus pada Bisnis	1. Berfokus pada eksplorasi data 2. Iteratif	1) Luas dan Komprehensif 2) Menekankan pada pengetahuan berguna
Kekurangan	1) Rigidity 2) Kurangnya Emphasis pada evaluasi berkelanjutan	1) Kurang fleksibel 2) Kurangnya fokus pada Bisnis	1) Kesulitan dalam implementasi 2) Waktu dan Sumber Daya
Tahapan	Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment	Sample, Explore, Modify, Model, Assess	Seleksi, Preprocessing, Transformasi dan Mining, Interpretasi/Evaluasi, Penggunaan.

Berdasarkan analisis pada Tabel 3.1 yang membandingkan kelebihan dan kekurangan berbagai metode, penelitian ini mengadopsi metode CRISP-DM sebagai pendekatan utama dalam proses Data Mining. Pemilihan metode ini didasarkan pada keunggulannya yang lebih menonjol dibandingkan metode lainnya, terutama dalam hal sistematisasi tahapan analisis yang lebih jelas dan mudah dipahami. CRISP-DM memungkinkan proses penelitian berjalan lebih efektif dan efisien dalam menghasilkan analisis data yang akurat. Model penelitian CRISP-DM merupakan metodologi standar dalam *data mining*. CRISP-DM terdiri dari enam tahapan utama:

1. *Business Understanding*

Memahami tujuan bisnis dan bagaimana analisis sentimen dapat memberikan wawasan bagi pengembang aplikasi Halo BCA.

2. *Data Understanding*

Mengumpulkan dan mengeksplorasi dataset ulasan dari Google Play Store untuk memahami karakteristiknya.

3. *Data Preparation*

Membersihkan data melalui *pre-processing* seperti penghapusan duplikasi, normalisasi teks, dan pembobotan fitur menggunakan TF-IDF.

4. *Modeling*

Menggunakan algoritma SVM, Naïve Bayes dan LSTM untuk klasifikasi sentimen ulasan.

5. *Evaluation*

Pengukuran performa model dilakukan dengan mengacu pada sejumlah indikator evaluasi seperti tingkat akurasi, nilai presisi, recall, serta F1-score untuk mengetahui tingkat keberhasilan metode klasifikasi yang digunakan.

## 6. Deployment

Hasil analisis akan didokumentasikan dan direkomendasikan untuk digunakan dalam pengambilan keputusan bisnis terkait peningkatan layanan aplikasi Halo BCA.

### 3.3 Teknik Pengumpulan Data

#### 3.3.1 Populasi dan Sampel

Seluruh ulasan pengguna aplikasi Halo BCA yang terdapat di Google Play Store menjadi populasi dalam penelitian ini sedangkan sampel diperoleh menggunakan pendekatan *purposive sampling*. Penelitian ini memanfaatkan kumpulan ulasan pengguna aplikasi Halo BCA yang dikumpulkan selama periode empat tahun terakhir, yakni dari 26 Juli 2021 hingga 26 Februari 2025, dengan total 17.499 ulasan. Pada penelitian ini jumlah datanya cukup besar sehingga menerapkan teknik *purposive sampling* untuk memilih sekitar 3.000 data sampel yang dianggap representatif terhadap populasi. Metode ini memungkinkan pemilihan sampel secara selektif berdasarkan kriteria tertentu, sehingga data yang dianalisis lebih relevan dan sesuai dengan tujuan penelitian.

Pengambilan sampel dilakukan dengan cara mengambil data dengan selisih setiap 2 baris dari dataset awal, sehingga sekitar 3.000 ulasan terpilih tetap mencerminkan distribusi keseluruhan data sebesar 17.499 ulasan. Pendekatan ini bertujuan untuk menjaga proporsionalitas representasi sentimen pengguna dan menghindari bias dalam pemilihan data. Setiap kategori sentimen dalam dataset memiliki peluang yang lebih seimbang untuk masuk ke dalam sampel penelitian.

Teknik *purposive sampling* digunakan karena pendekatan ini memungkinkan pemilihan data yang lebih terfokus dan relevan dengan tujuan analisis sentimen. Dalam konteks penelitian ini, metode ini memastikan bahwa sampel yang diambil mencerminkan variasi opini pengguna terhadap aplikasi Halo BCA, baik dalam sentimen positif ataupun negatif.

### 3.3.2 Periode Pengambilan Data

Data ulasan aplikasi Halo BCA dikumpulkan dalam rentang waktu tertentu untuk memastikan keakuratan dan relevansi analisis sentimen. Periode pengambilan data dilakukan selama 4 tahun terakhir sebelum penelitian dimulai, yaitu dari 26 Juli 2021 – 26 Februari 2025.

## 3.4 Variabel Penelitian

Penelitian ini melibatkan dua jenis variabel utama, yaitu variabel independen dan variabel dependen, yang masing-masing berperan dalam menganalisis hubungan serta pengaruh dalam penelitian:

### 3.4.1 Variabel Independen (X)

1. Teks Ulasan Pengguna

Ulasan yang diberikan oleh pengguna aplikasi Halo BCA di Google Play Store yang mencerminkan opini mereka terhadap aplikasi.

2. Fitur Teks yang diproses

Hasil dari *pre-processing* data seperti tokenisasi, stemming, *stopword removal*, *labelling* dan pembobotan menggunakan metode TF-IDF.

3. Algoritma yang digunakan

Metode klasifikasi SVM, Naïve Bayes dan LSTM digunakan untuk menganalisis sentimen ulasan pengguna.

### 3.4.2 Variabel Dependen (Y)

1. Kategori Sentimen

Klasifikasi sentimen ulasan pengguna yang terdiri dari positif, dan negatif berdasarkan hasil analisis dari model yang digunakan.

2. Akurasi Model

Tingkat keberhasilan model dalam melakukan klasifikasi yang diukur dengan metrik akurasi, presisi, recall, dan F1-score.

Penentuan variabel ini bertujuan untuk memastikan bahwa penelitian memiliki fokus yang jelas dalam menganalisis sentimen ulasan pengguna

aplikasi Halo BCA serta mengevaluasi performa metode yang digunakan dalam klasifikasi sentimen.

### 3.5 Teknik Analisis Data

Penelitian ini menerapkan beberapa tahapan analisis data yang bertujuan untuk mengolah dan mengevaluasi hasil klasifikasi sentimen terhadap ulasan pengguna. Prosedur analisis data yang digunakan dirancang secara sistematis untuk mendukung tujuan penelitian:

#### 1. Preprocessing Data

Pengambilan data dilakukan melalui Google Play Store dengan menerapkan teknik *web scraping*. Selanjutnya, data yang diperoleh dibersihkan melalui serangkaian proses, termasuk *case folding*, tokenisasi, penghapusan *stopword*, *stemming*, *labelling* serta pembobotan menggunakan TF-IDF, Tokenisasi dan *Padding*. Pada metode berbasis pembelajaran mesin konvensional seperti Naïve Bayes dan SVM, representasi teks menggunakan TF-IDF efektif karena menghasilkan vektor fitur berdimensi tetap yang mewakili frekuensi penting kata dalam dokumen secara numerik. Namun, untuk model *deep learning* seperti LSTM, pendekatan TF-IDF kurang sesuai karena LSTM membutuhkan urutan kata yang mempertahankan konteks dan hubungan antar kata secara berurutan sehingga untuk LSTM dilakukan proses tokenisasi yang mengubah setiap kata menjadi indeks numerik berdasarkan kamus kata, lalu dilakukan *padding* agar setiap input memiliki panjang yang sama. *Padding* ini penting untuk memastikan semua data input memiliki ukuran seragam sehingga dapat diproses dalam *batch* selama pelatihan LSTM. LSTM dapat mempelajari pola dan dependensi sekuensial dalam teks secara efektif, yang berbeda dengan TF-IDF yang menghilangkan urutan kata. Rangkaian proses ini dilakukan untuk meningkatkan mutu data sehingga lebih layak untuk dianalisis secara mendalam.

## 2. Pemodelan dan Implementasi

Data yang telah melalui tahap pemrosesan akan dimanfaatkan sebagai *input* dalam pemodelan klasifikasi sentimen menggunakan tiga metode, yaitu SVM, Naïve Bayes, dan LSTM. Proses *training* model dilakukan dengan bantuan pustaka Scikit-learn dalam bahasa pemrograman Python untuk mengklasifikasikan ulasan menjadi dua kelompok sentimen, yakni positif dan negatif.

## 3. Evaluasi Model

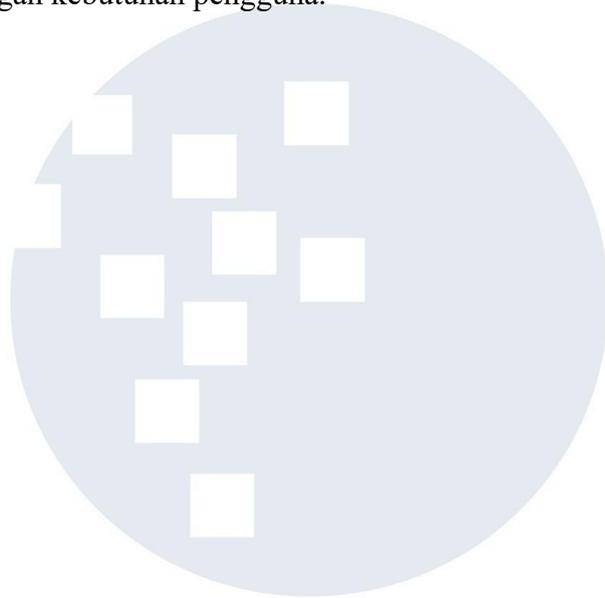
Model yang dikembangkan kemudian dianalisis dengan metrik seperti akurasi, presisi, *recall*, dan F1-score sebagai tolak ukur performa klasifikasi sentimen. Hasil evaluasi ini akan menjadi dasar dalam menentukan metode yang paling efektif dalam mengelompokkan ulasan pengguna aplikasi Halo BCA secara akurat.

## 4. Interpretasi Hasil

Tahapan interpretasi hasil dilakukan untuk memahami secara lebih mendalam pola sentimen yang terkandung dalam ulasan pengguna terhadap aplikasi Halo BCA. Analisis ini tidak hanya berfokus pada angka-angka evaluasi model tetapi juga berupaya mengungkap kecenderungan sentimen yang dominan, baik positif maupun negatif dari perspektif pengguna dan untuk memperkuat pemahaman terhadap hasil klasifikasi maka dibuat beberapa visualisasi seperti grafik distribusi sentimen, grafik frekuensi kata (*Word Cloud*) dan tabel perbandingan performa model. Melalui pendekatan ini, hasil evaluasi model tidak hanya disajikan dalam bentuk metrik numerik tetapi juga didukung oleh data visual yang membantu penjelasan bagaimana model membaca pola dalam teks ulasan. Analisis ini menjadi dasar dalam memahami persepsi pengguna serta mengidentifikasi area perbaikan pada layanan aplikasi berdasarkan umpan balik yang diberikan.

Melalui proses analisis ini, penelitian berupaya memberikan pemahaman yang lebih mendalam mengenai sentimen pengguna terhadap aplikasi Halo BCA melalui analisis data ulasan dari Google Play Store. Data diproses melalui

tahap *pre-processing* seperti tokenisasi, TF-IDF, dan *padding* yang disesuaikan dengan model klasifikasi Naïve Bayes, SVM, dan LSTM. Performa setiap metode dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan F1-score. Hasil klasifikasi kemudian dianalisis dan divisualisasikan untuk mengidentifikasi pola sentimen pengguna. Temuan ini diharapkan menjadi dasar dalam merumuskan strategi pengembangan layanan yang lebih optimal dan sesuai dengan kebutuhan pengguna.



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA