

BAB III

METODOLOGI DAN PERANCANGAN SISTEM

3.1 Objek Penelitian

Dalam penelitian ini, akan dilakukan sentimen analisis dengan objek penelitian yaitu opini pengguna aplikasi Bibit yang ada di *Google Play Store*. Dari sentimen analisis tersebut bisa didapatkan berapa banyak jumlah komentar positif, netral serta negatif mengenai aplikasi tersebut sehingga bisa ditarik kesimpulan berdasarkan jumlah dari analisis tersebut bagaimana pendapat mengenai pemakaian aplikasi Bibit dari para penggunanya khususnya aplikasi yang berbasis Android.

3.2 Variabel Penelitian

Dalam penelitian ini, variabel yang akan digunakan terbagi menjadi dua jenis yaitu:

a. Variabel Independen

Variabel Independen pada penelitian ini adalah opini dari para pengguna aplikasi Bibit yang diambil dari *Google Play Store*

b. Variable Dependen

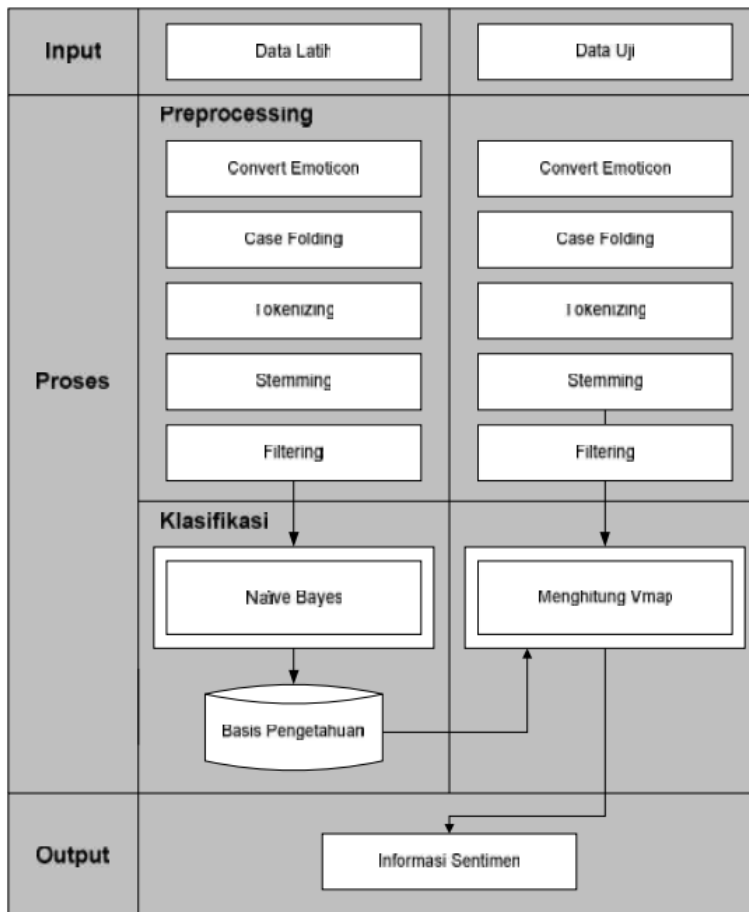
Variabel dependen pada penelitian ini adalah hasil kelas klasifikasi yaitu nilai positif, netral dan negatifnya.

3.3 Metode Penelitian

Algoritma yang akan dilakukan dalam penelitian ini dengan algoritma, *Support Vector Machine*, *Random Forest*, *Naïve Bayes*. Untuk data diambil dari *review* terhadap aplikasi Bibit di *Google Play Store*.

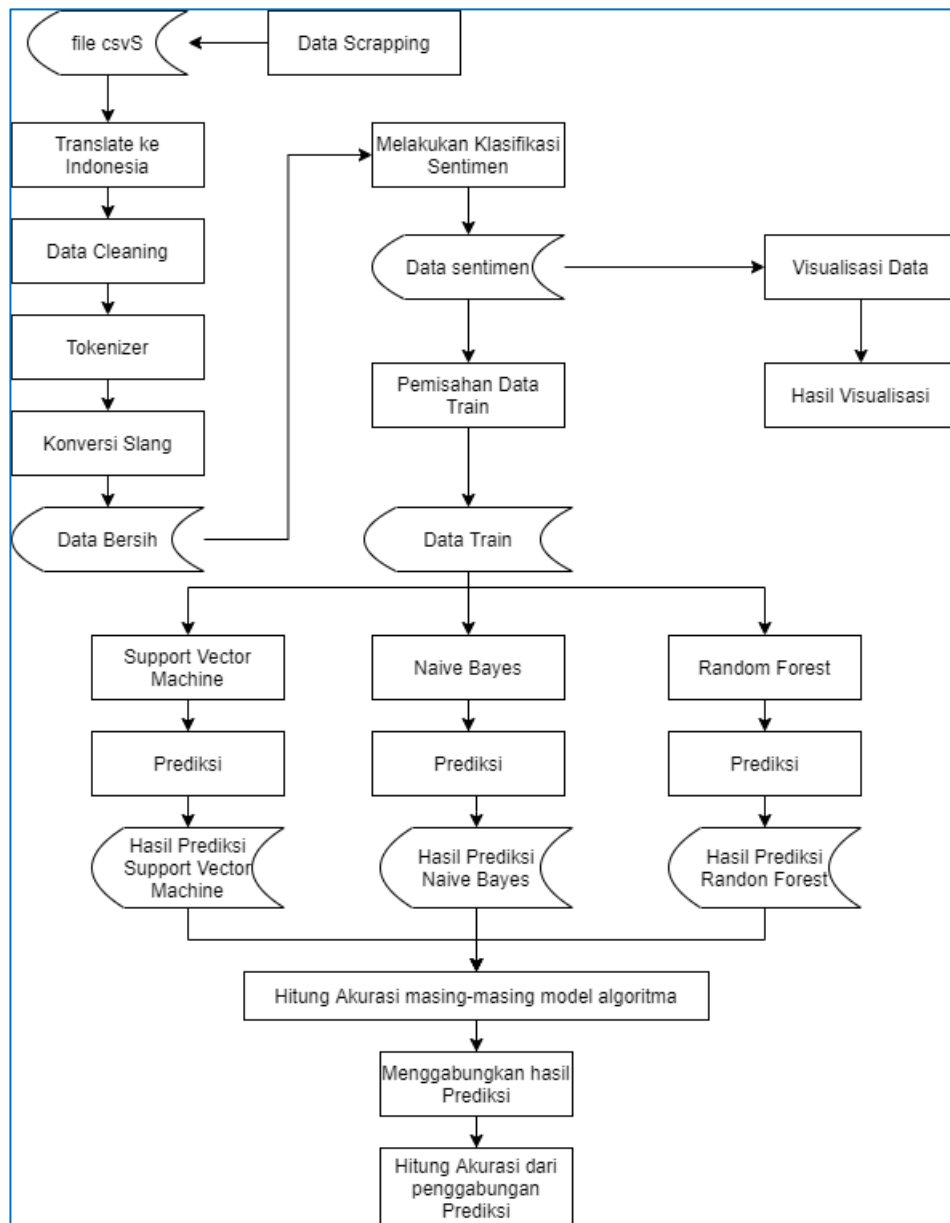
3.3.1 Alur Penelitian

Penelitian ini akan menggunakan alur penelitian yang digunakan oleh Lorosae dan Prakoso sebagai referensi dengan judul “Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan *Naïve Bayes*” [28] dengan melakukan perubahan dan penambahan sesuai dengan tujuan pada penelitian ini. Berikut adalah gambar alur penelitian sebelum dilakukannya modifikasi:



Gambar 3. 1 alur penelitian yang digunakan sebelum dimodifikasi

Alur penelitian yang telah dimodifikasi dan akan digunakan pada penelitian ini adalah sebagai berikut.



Gambar 3. 2 Alur Penelitian yang telah dimodifikasi

3.3.2 Fase persiapan Data

Pada fase ini yang dilakukan adalah persiapan data source atau dataset yang akan digunakan untuk mengukur analisis sentimen. Pada pengambilan data source yang dilakukan pada *review* aplikasi Bibit di *Google Play Store* menggunakan

google-play-scraper yang berguna untuk pengambilan comments yang ada di post pada instagram. Tetapi tidak semua *review* bisa diambil, dan didapatkan *review* sejumlah 10.468 *review* terhadap aplikasi Bibit.

3.3.3 *Preprocessing Data*

Tahap ini di berlakukannya proses pengolahan data dari data *source* yang telah diambil, agar dapat digunakan dalam model klasifikasi dengan cara melakukan pembersihan data agar tidak terdapat error dan data tidak konsisten [29]. Tahap ini terdiri dari beberapa proses sebagai berikut :

1. *Translate Data*

Tahap ini mengubah semua data yang sebelumnya menggunakan Bahasa Inggris menjadi Bahasa Indonesia untuk mendapatkan hasil yang sama dan konsisten.

2. *Cleansing Data*

Tahap ini diberlakukan proses pembersihan data agar data yang akan digunakan tidak memiliki error [30], maupun ketidak konsistenan. Pada proses menggunakan aplikasi R Studio untuk melakukan pembersihan data. Data yang dibersihkan seperti huruf-huruf hasil translate

yang tidak terbaca sebelumnya seperti “ðÿ~” yang ditemukan dalam data

3. *Tokenize Data*

Fase ini diberlakukannya perubahan dari suatu kalimat menjadi kata-kata dengan konsep setiap ada spasi disuatu kalimat, akan dipecah menjadi kata-kata yang nanti akan mempermudah pengerjaan pada proses selanjutnya.

4. Konversi *Slangword* dan *Stemming*

Fase ini diberlakukannya perubahan kata-kata yang tidak baku menjadi kata baku dan mengubah kata-kata yang sering disingkat menjadi kata baku dari kata tersebut [30] seperti contohnya mengubah kata “lgi”, “ngk” menjadi kata “lagi” dan “tidak”. Proses ini menggunakan kamus dari *website* [31] yang menjadi acuan untuk mengubah kata-kata menjadi baku pada kalimat yang ada.

5. Penghapusan *Stop Words*

Fase ini diberlakukannya penghapusan kata penghubung seperti “dan”,

“atau”, “yang”, serta kata penghubung lainnya, sehingga saat melakukan analisis sentimen akan lebih mudah dan mencari kata yang sering muncul.

3.3.4 Problem Solving

Berikut adalah kelebihan dan kekurangan dari tiap jenis algoritma yang akan digunakan dalam penelitian :

Tabel 3. 1 Tabel kelebihan dan kekurangan untuk setiap algoritma

Algoritma	Kelebihan	Kekurangan	Fungsi
<i>Random Forest</i> [32]	<ol style="list-style-type: none"> 1. Serbaguna 2. hyperparameter default yang digunakannya sering kali menghasilkan hasil prediksi yang baik. Memahami hyperparameter cukup mudah, dan jumlahnya juga tidak banyak. 	<p>banyaknya pohon dapat membuat algoritme terlalu lambat dan tidak efektif untuk prediksi waktu nyata.</p>	<p>- Klasifikasi</p>
<i>Support Vector Machine</i> [33]	<ol style="list-style-type: none"> 1. memiliki akurasi yang tinggi 2. Berfungsi dengan baik pada set data yang lebih kecil dan lebih bersih 3. Ini bisa lebih efisien karena menggunakan subset poin pelatihan 	<ol style="list-style-type: none"> 1. Tidak cocok untuk kumpulan data yang lebih besar karena waktu pelatihan dengan <i>SVM</i> bisa tinggi 2. Kurang efektif pada set data yang 	<p>- Klasifikasi - Regresi</p>

		lebih rumit dengan kelas yang tumpang tindih	
<i>Naïve Bayes</i> [34]	<ol style="list-style-type: none"> 1. Meskipun asumsi naif jarang benar, algoritme bekerja dengan sangat baik dalam banyak kasus 2. Menangani data berdimensi tinggi dengan baik. Mudah untuk memparalelkan dan menangani data besar dengan baik 3. Berkinerja lebih baik daripada model yang lebih rumit saat kumpulan datanya kecil 	<ol style="list-style-type: none"> 1. Estimasi probabilitas seringkali tidak akurat karena asumsi yang naif. Tidak ideal untuk penggunaan regresi atau estimasi probabilitas 2. Jika datanya berlimpah, model lain yang lebih rumit cenderung mengungguli <i>Naïve Bayes</i> 	- Klasifikasi

3.3.5 Visualisasi

Pada fase ini diberlakukannya visualisasi data menurut hasil yang ada dan visualisasi data berupa *graphing* yang mudah dipahami

3.4 Teknik Pengumpulan Data

Teknik yang digunakan untuk pengambilan data untuk penelitian ini adalah dengan menggunakan *google-play-scraper* yang didapatkan dari *website*. Data tersebut diperoleh berdasarkan komen yang dilakukan

oleh pengguna aplikasi Bibit terhadap 10468 *review* yang dilakukan pada aplikasi Bibit di *Google Play Store* pada tanggal, yaitu 10 Januari 2019 hingga 27 Februari 2021. 10 Januari 2019 merupakan tanggal dimana aplikasi Bibit pertama kali dibuka dan 27 Februari adalah tanggal pengambilan data. 10468 *review* didapatkan sudah cukup dikarenakan *review* tersebut memiliki jumlah lebih banyak daripada penelitian terdahulu Lorosae dan Prakoso yaitu sebanyak 2000 kalimat [28].