

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Objek Penelitian**

Objek penelitian yang digunakan adalah *review* pengguna produk kecantikan Wardah Sunscreen Gel SPF 30PA++ pada website *femaledaily*.

#### **3.2. Subjek Penelitian**

Subjek penelitian yang digunakan adalah pengguna produk kecantikan Wardah Sunscreen Gel SPF 30PA++.

#### **3.3. Alat dan Bahan Penelitian**

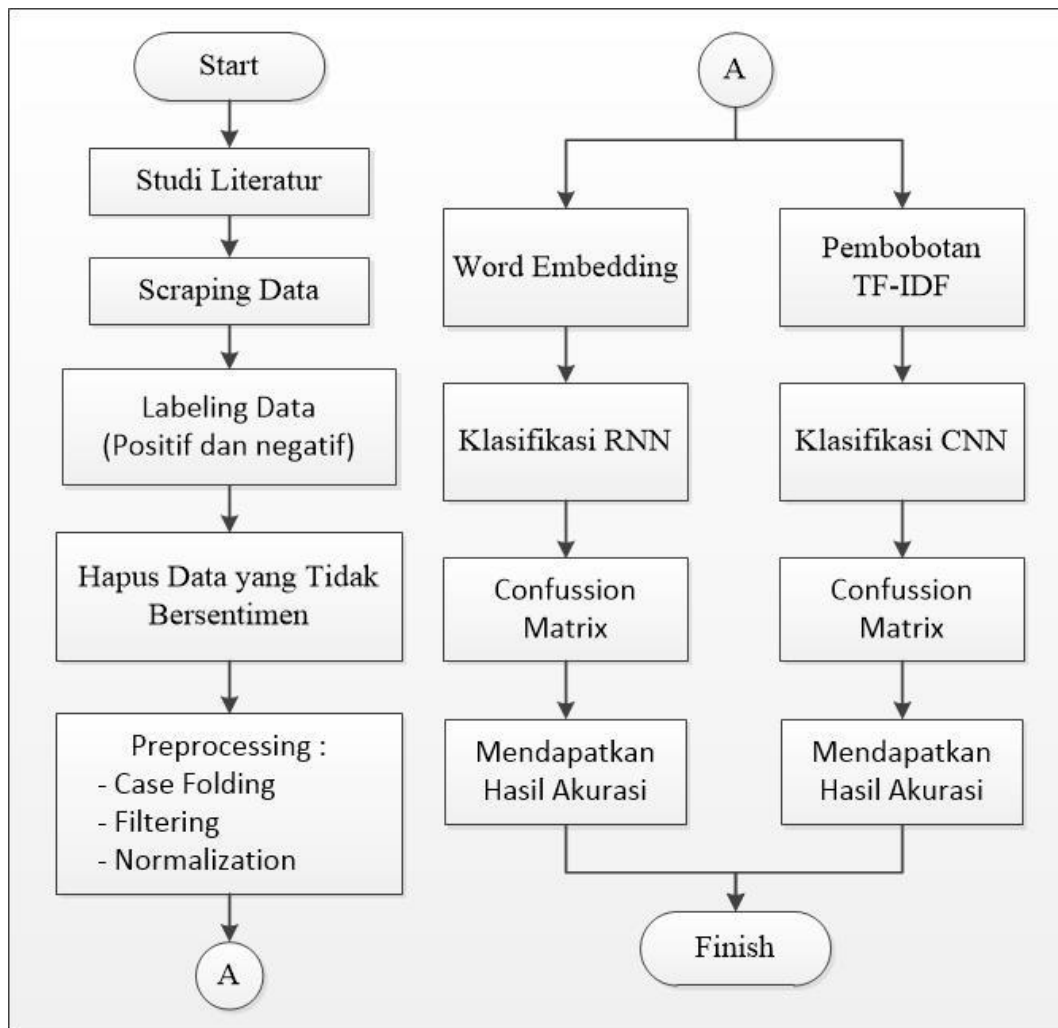
**Alat:**

Notebook dengan spesifikasi :  
Core i5 8<sup>th</sup> gen  
Graphics card GTX 1050  
Ram 8 GB  
Memory SSD 256GB

**Bahan:**

*Review* pengguna produk  
Wardah pada komentar produk  
laman *Femaledaily*.

### 3.4. Metodologi Penelitian



**Gambar 3. 1 Diagram Alir Penelitian**

Gambar 3.1 merupakan gambaran flowchart metodologi penelitian yang akan dilakukan pada penelitian kali ini. Metode pada penelitian ini disusun secara sistematis sehingga penelitian bersifat logis serta tetap mengikuti langkah-langkah ilmiah yang berlaku. Penyusunan penelitian yang sistematis juga bertujuan untuk mempermudah peneliti untuk melakukan penelitian.

Metode yang digunakan pada penelitian ini yaitu menggunakan metode RNN dan CNN untuk melakukan analisis sentimen. Namun, sebelum melakukan proses *modelling*, ada beberapa tahapan yang harus dilakukan terlebih dahulu untuk mendapatkan akurasi sesuai dengan keinginan peneliti. Metode yang digunakan dipilih melalui fase perbandingan beberapa metode yang dapat

digunakan untuk melakukan analisis sentimen, RNN dan CNN terpilih sebagai metode yang akan digunakan pada penelitian ini dengan alasan bahwa metode tersebut memiliki kelebihan dalam pengolahan data yang kompleks serta dapat memahami pola yang memiliki tingkat kompleksitas lebih tinggi untuk dapat dipahami dengan menggunakan metode lainnya.

### **3.3.1. Scraping Data**

Jenis data yang digunakan dalam penelitian dibagi menjadi data primer dan sekunder. Data yang digunakan pada penelitian ini menggunakan data sekunder yang diambil dari laman *femaledaily*. Pengambilan data menggunakan teknik scraping dengan menggunakan library *Beautifulsoup* yang disediakan python dengan mengambil data dari HTML pada laman yang terkait. Pengumpulan data dilakukan dari komentar paling hingga komentar terakhir pada tanggal 22 Juni 2022.

### **3.3.2. Labeling Data dan Penghapusan Data Tidak Bersentimen**

Data yang sudah diperoleh, selanjutnya akan dilakukan labeling data dengan tujuan untuk melakukan pemilahan data dengan sentimen negatif atau positif. Hal ini ditujukan agar sistem dapat memahami makna dari setiap kata yang akan diujikan. Labeling dilakukan secara manual oleh peneliti dengan tetap mematuhi aturan labeling secara umum. Dataset yang telah melalui proses labeling data, diharapkan berjumlah seimbang agar mendapatkan akurasi yang sesuai dengan keinginan penulis. Setelah data yang melalui proses labeling jumlahnya seimbang, kemudian data yang tidak melalui proses labeling atau maknanya tidak masuk ke sentiment positif ataupun negatif akan dihapus.

### **3.3.3. Preprocessing**

Data yang sudah melalui proses labeling, kemudian akan dilanjutkan pada proses *preprocessing*. Pada tahap ini akan dilakukan pembersihan lebih lanjut terhadap data yang dimiliki hingga menjadi data yang siap untuk diolah. Tahap *preprocessing* memiliki beberapa proses sebagai berikut.

### 1. Case Folding

Pada proses ini akan dilakukan perubahan huruf yang awalnya besar (uppercase), menjadi huruf kecil (lowercase) dengan tujuan untuk membuat data menjadi seragam. Berikut merupakan gambaran dari proses case folding dengan contoh komentar dari laman *femaledaily* pada tabel 3.1.

**Tabel 3.1 Contoh Penerapan Proses Case Folding**

Input	Output
Sunscreen pertama yg ku coba, pake ini pas masih jaman sma. Dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah. Utk produk wardah bagus2 cuma mungkin yg ini yg kurang cocok di aku..	sunscreen pertama yg ku coba, pake ini pas masih jaman sma. dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah. utk produk wardah bagus2 cuma mungkin yg ini yg kurang cocok di aku..

### 2. Filtering

Pada proses ini dilakukan penyaringan kata untuk menghilangkan simbol-simbol, karakter khusus, serta emoticon yang terdapat pada suatu kalimat. Proses ini bertujuan agar pengolahan tingkat error tidak tinggi dikarenakan data yang tidak penting. Contoh data yang akan dihapus yaitu koma (,), titik (.), hashtag (#), bintang (\*), dan banyak karakter khusus serta simbol lainnya. Berikut merupakan gambaran dari proses varian node buffering dengan contoh komentar dari laman *femaledaily*. Berikut merupakan gambaran dari proses case folding dengan contoh komentar dari laman *femaledaily* pada tabel 3.2.

**Tabel 3.2 Contoh Penerapan Proses Normalisasi**

Input	Output
sunscreen pertama yg ku coba, pake ini pas masih jaman sma. dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah. utk produk wardah bagus2	sunscreen pertama yg ku coba pake ini pas masih jaman sma dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah utk produk wardah bagus cuma mungkin yg ini yg

cuma mungkin yg ini yg kurang cocok di aku..	kurang cocok di aku
--	---------------------

### 3. Normalization

Pada proses ini dilakukan perubahan kalimat, apabila ada kata yang bersifat tidak baku atau kata slang diubah menjadi kata bakunya serta merubah kata yang disingkat agar menjadi aslinya. Contohnya seperti “gue” diubah menjadi “saya”, “lo” diubah menjadi “anda”, serta beberapa kata slang lain. Berikut merupakan gambaran dari proses normalisasi dengan contoh komentar dari laman *femaledaily*. Berikut merupakan gambaran dari proses normalisasi dengan contoh komentar dari laman *femaledaily* pada tabel 3.3.

**Tabel 3.3 Contoh Penerapan Proses Filtering**

Input	Output
<p>sunscreen pertama yg ku coba pake ini pas masih jaman sma dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah utk produk wardah bagus cuma mungkin yg ini yg kurang cocok di aku</p>	<p>sunscreen pertama yang aku coba pakai ini pas masih jaman sma dikulitku kurang cocok sama sunscreen ini buat wajah jadi beruntusan dan ngerasa setelah pakai ini jadi kelihatan kusam wajah untuk produk wardah bagus cuma mungkin yang ini yang kurang cocok di aku</p>

Berikut merupakan contoh proses *preprocessing* dari awal hingga akhir.

**Tabel 3.4 Contoh *Preprocessing* dengan Data Sampel**

Proses	Output
Case Folding	sunscreen pertama yg ku coba, pake ini pas masih jaman sma. dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah. utk produk wardah bagus2 cuma mungkin yg ini yg kurang cocok di aku..
Filtering	sunscreen pertama yg ku coba pake ini pas masih jaman sma. dikulitku kurang cocok sm sunscreen ini buat wajah jd beruntusan dan ngerasa setelah pake ini jd keliatan kusam wajah utk produk wardah bagus cuma mungkin yg ini yg kurang cocok di aku
Normalization	sunscreen pertama yang aku coba pakai ini pas masih jaman sma dikulitku kurang cocok sama sunscreen ini buat wajah jadi beruntusan dan ngerasa setelah pakai ini jadi kelihatan kusam wajah untuk produk wardah bagus cuma mungkin yang ini yang kurang cocok di aku

#### 3.3.4. Pembobotan TF-IDF

Pada tahap ini dilakukan pembobotan TF-IDF untuk mengetahui seberapa penting suatu kata dalam dokumen yang diolah. Konsep yang digunakan pada proses yaitu melihat frekuensi kemunculan kata dalam dokumen serta invers dari nilai frekuensi dokumen yang berisikan kata tersebut. TF-IDF dilakukan dengan menggunakan library yang terdapat pada python yaitu menggunakan `sklearn.feature_extraction.text.TfidfVectorizer` untuk memanggil proses TF-IDF yang ada dalam library tersebut. Berikut contoh pembobotan TF-IDF menggunakan data sampel.

**Tabel 3.5 Contoh Dataset yang Akan Dilakukan Pembobotan TF-IDF**

Username	Komentar
firdalatas	kok saya pakai ini tetap gosong yaa apa cuma aku atau yang lain juga jadi ceritanya lagi liburan beberapa hari begitu dipantai dan aku ngerasa udah ngehindar dari sengatan matahari walaupun udah ngerasa pakai sunscreen ini tapi hasilnya tetap gosong
sandrandhini	sunscreen pertama yang aku coba pakai ini pas masih jaman sma dikulitku kurang cocok sama sunscreen ini buat wajah jadi beruntusan dan ngerasa setelah pakai ini jadi kelihatan kusam wajah untuk produk wardah bagus cuma mungkin yang ini yang kurang cocok di aku

Setelah data melalui tahap *preprocessing*, kemudian akan dilanjutkan dengan proses pembobotan TF-IDF. Berikut contoh perhitungan TF-IDF dengan menggunakan contoh dataset pada tabel 3.5 pada kata kok, maka perhitungannya sebagai berikut.

1. Kata 'kok' muncul hanya pada dokumen ke-1 maka dari itu  $D1 = 1$  dan  $D2 = 0$ .
2. Melakukan perhitungan  $df_t$  yang merupakan jumlah banyaknya dokumen yang mengandung kata 'kok'. Kemunculan kok hanya pada D1 maka dari itu DF nya 1.
3. Melakukan perhitungan IDF dengan rumus  $\log \frac{N}{df_t}$ , N disini merupakan jumlah total dokumen yang di gunakan. Maka hasil yang didapatkan  $\log(2/1) = 0,301$ .
4. Melakukan perhitungan bobot akhir  $W_{t,d}$  dengan rumus  $tf_{t,d} \times idf_t$ , maka hasilnya  $1 \times 0,301 = 0,301$

**Tabel 3.6 Tahap Perhitungan TF-IDF**

Term(t)	$tf_{t,d}$		$df_t$	$idf_t$	$W_{t,d}$	
	D1	D2			D1	D2
kok	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
saya	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
pakai	2	2	2	$\text{Log}(2/2) = 0$	0	0
ini	2	4	2	$\text{Log}(2/2) = 0$	0	0
tetap	2	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
gosong	2	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Apa	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Cuma	1	1	2	$\text{Log}(2/2) = 0$	0	0
Atau	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Yang	1	3	2	$\text{Log}(2/2) = 0$	0	0
Lain	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Juga	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Jadi	1	2	2	$\text{Log}(2/2) = 0$	0	0
Ceritanya	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0



Term(t)	$tf_{t,d}$		$df_t$	$idf_t$	$W_{t,d}$	
	D1	D2			D1	D2
				0,301		
Lagi	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Liburan	1	0	2	$\text{Log}(2/2) = 0$	0	0
Hari	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
gitu	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Dipantai	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Dan	1	1	2	$\text{Log}(2/2) = 0$	0	0
ngerasa	2	1	2	$\text{Log}(2/2) = 0$	0	0
Uda	2	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Ngehinda r	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Dari	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Sengatan	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Matahari	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
walaupun	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Sunscreen	1	2	2	$\text{Log}(2/2) = 0$	0	0

Term(t)	$tf_{t,d}$		$df_t$	$idf_t$	$W_{t,d}$	
	D1	D2			D1	D2
				0		
Tapi	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Hasilnya	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Gosong	1	0	1	$\text{Log}(2/1) = 0,301$	0,301	0
Pertama	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Coba	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Pas	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Masih	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Jaman	0	1	2	$\text{Log}(2/2) = 0$	0	0
Sma	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Dikulitku	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Kurang	0	2	1	$\text{Log}(2/1) = 0,301$	0	0,602
Cocok	0	2	1	$\text{Log}(2/1) = 0,301$	0	0,602
Buat	1	1	2	$\text{Log}(2/2) = 0$	0	0
wajah	0	2	1	$\text{Log}(2/1) = 0,301$	0	0,602

Term(t)	$tf_{t,d}$		$df_t$	$idf_t$	$W_{t,d}$	
	D1	D2			D1	D2
				0,301		
beruntusan	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Setelah	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
Kelihatan	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
kusam	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
untuk	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
produk	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
wardah	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
bagus	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
mungkin	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301
di	0	1	1	$\text{Log}(2/1) = 0,301$	0	0,301

Keterangan :

D1, D2 = Dokumen ke 1 dan 2

$df_t$  = Jumlah dokumen yang mengandung kata yang sedang dihitung

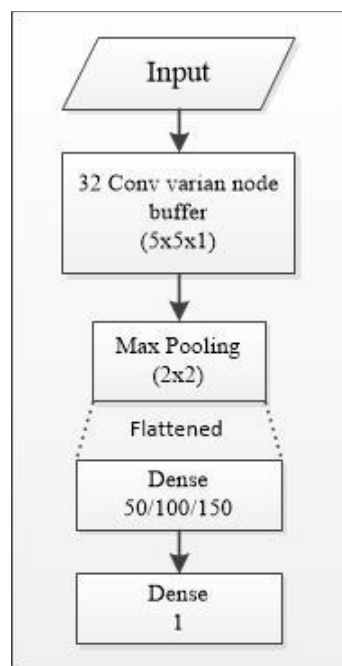
$idf_t$  = Jumlah invers frekuensi dokumen tiap kata

$W_{t,d}$  = Pembobotan TF-IDF

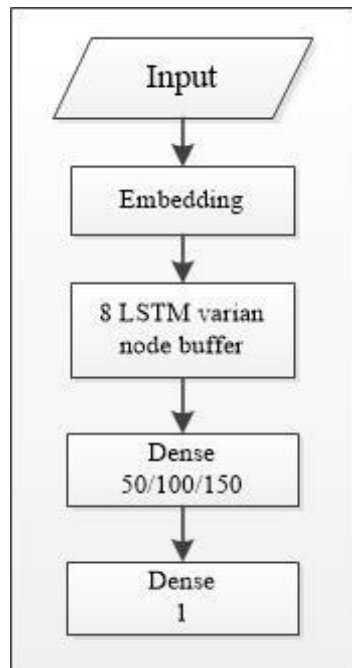
### 3.3.5. Pengujian Model

Proses analisis data pada penelitian ini dilakukan dengan menggunakan algoritma RNN dan CNN berbasis python dengan tools yang digunakan pada penelitian ini yaitu *Google Colab*. Data yang sudah dilakukan *preprocessing* dan pembobotan selanjutnya akan dilakukan analisis untuk dapat mengetahui sentimen dari suatu kalimat. Tingkat akurasi yang dihasilkan dari proses ini, sangat dipengaruhi dengan proses *preprocessing*, apabila proses tersebut tidak dilakukan dengan benar maka tingkat akurasi yang dihasilkan akan terpengaruh.

Proses pengujian model dilakukan dengan membagi data menjadi dua kategori yaitu sebagai data *training* (pelatihan) serta data *testing* (pengujian). Data yang sudah melalui proses *preprocessing* dan pembobotan akan dibagi kelompok berdasarkan pembagian *training* dan *testing*nya. Terdapat 3 kelompok yang akan digunakan yaitu kelompok pertama menggunakan perbandingan 8:2, kelompok kedua menggunakan perbandingan 7:3, dan kelompok ketiga menggunakan perbandingan 9:1. Hal ini bertujuan untuk membandingkan hasil dari masing-masing kelompok, manakah perbandingan yang memiliki hasil akurasi serta waktu *modelling* terbaik. Berikut merupakan diagram alir metode RNN dan CNN.



**Gambar 3. 2 Diagram Alir Metode CNN**



**Gambar 3. 3 Diagram Alir Metode RNN**

### **3.3.6. Confusion Matrix**

Setelah proses pengujian model dilakukan, kemudian akan dilanjutkan proses validasi serta evaluasi untuk mendapatkan nilai performa model serta meningkatkan performa model yang digunakan. Validasi model ditujukan untuk menampilkan nilai akurasi dari uji model yang sudah dilakukan sementara evaluasi dilakukan untuk mengetahui kesalahan yang terjadi pada model untuk meningkatkan tingkat akurasi yang didapatkan.

Proses validasi dan evaluasi model pada penelitian ini menggunakan confusion matrix dengan hasil akhir berfokus pada tingkat akurasi yang dihasilkan. Penggunaan confusion matrix pada proses ini dengan menggunakan library yang telah disediakan dari jupyter notebook lebih tepatnya sklearn. Library yang dirujuk akan dipanggil dengan menggunakan syntax `sklearn.metrics.confusion_matrix`.