

BAB III

METODE PENELITIAN

3.1 Subjek dan objek penelitian

Subjek penelitian ialah benda, orang, atau tempat yang diamati. Opini dari *user* media sosial *Twitter* di Indonesia dari sebuah kasus *bullying* yang dilakukan oleh Idol K-pop Kim Garam menjadi subjek dalam penelitian ini. Objek merupakan atribut dari kegiatan yang ditetapkan oleh peneliti. Objek dari penelitian ini yaitu perbandingan dari dua metode *naive bayes classifier* dan *support vector machine* dalam klasifikasi opini kasus *bullying* yang dilakukan oleh Idol K-pop Kim Garam.

3.2 Alat dan Bahan

3.2.1 Perangkat Keras

Pada penelitian ini menggunakan *hardware* yaitu laptop dengan spesifikasi sebagai berikut:

1. *Processor Intel(R) Core(TM) i5-8265U*
2. RAM 4 GB
3. Storage 1TB
4. Display 14” Full HD
5. Graphics Intel UHD Graphics 620 dan Nvidia GeForce MX23
VRAM 2GB GDDR5

3.2.2 Perangkat Lunak

Beberapa *software* yang dipakai pada penelitian ini yaitu sebagai berikut:

1. Sistem operasi windows 11
2. *Google colaboratory*

3. *Command prompt*

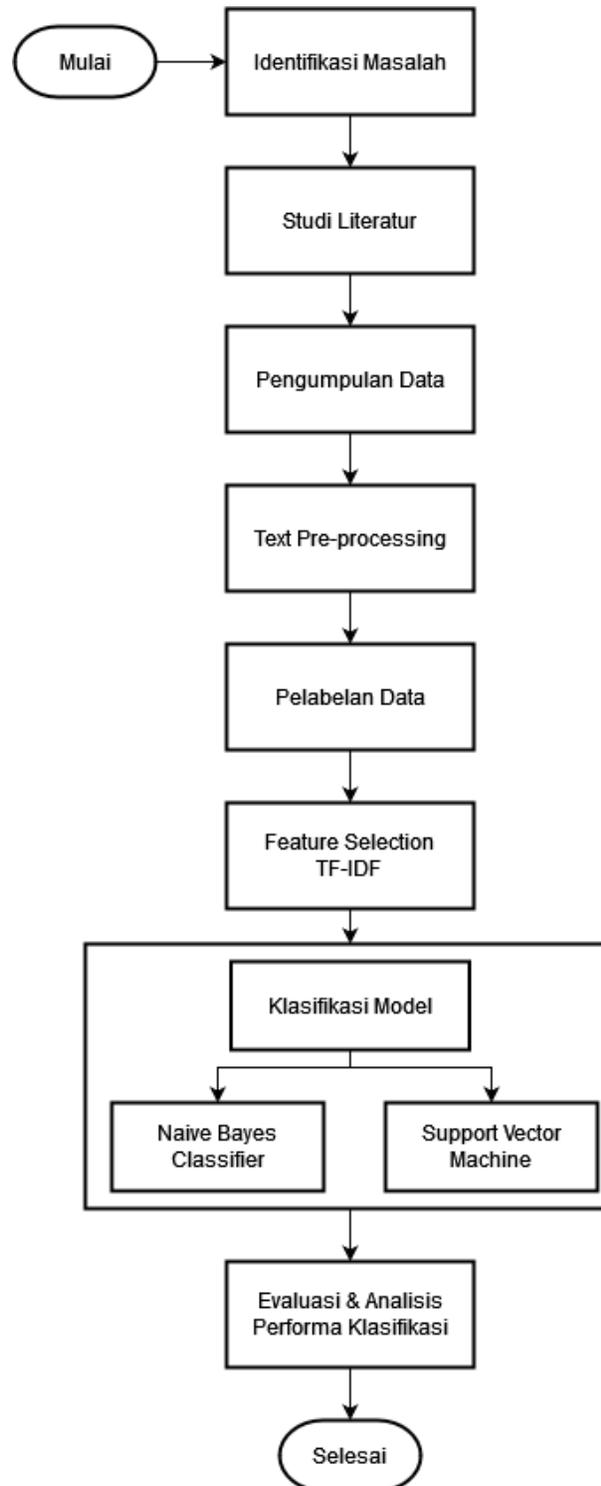
4. *Browser*

3.2.3 Bahan

Penelitian ini menggunakan bahan berupa kumpulan data *tweet* dari media sosial *Twitter* dalam kurun waktu 22 April 2022 hingga 10 Januari 2023 tentang rumor kasus *bullying* oleh Kim Garam yang selanjutnya data dikategorikan ke dalam tiga kelas yaitu positif, negatif dan netral.

3.3 Diagram alir penelitian

Beberapa tahapan yang dilakukan untuk menyelesaikan penelitian “Perbandingan Metode *Naive Bayes Classifier* dan *Support Vector Machine* untuk Klasifikasi Respon Kasus *Bullying* pada *Twitter*” yang diilustrasikan pada diagram alir seperti Gambar 3.1.



Gambar 3. 1 Diagram alir penelitian

3.3.1 Identifikasi Masalah

Langkah awal dari penelitian ini yaitu mengidentifikasi dan merumuskan permasalahan apa yang akan diteliti. Kemudian tentukan topik dan metode apa yang akan digunakan pada penelitian tersebut. Topik bisa ditentukan dengan melihat masalah yang sedang terjadi atau ramai dibicarakan saat ini. Selain itu penentuan topik juga dapat dilihat dari penelitian terdahulu yang sudah ada, namun ada perbedaan seperti pada metode yang digunakan.

3.3.2 Studi Literatur

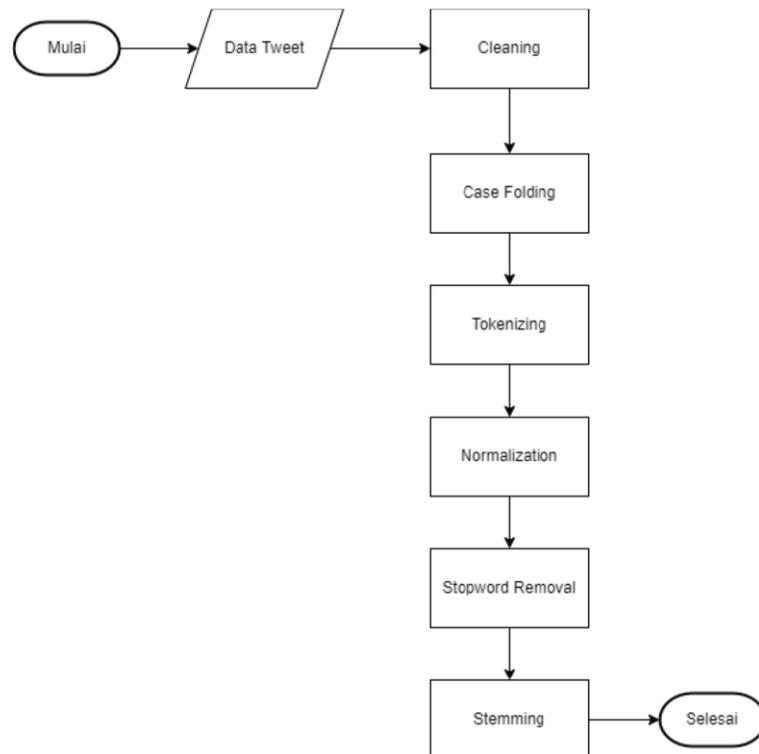
Tahap ini dilakukan sebagai riset kepustakaan. Penelitian dilakukan dengan mengumpulkan beberapa data terkait dengan topik permasalahan pada penelitian yaitu klasifikasi *bullying* Idol K-Pop, dan penelitian ini menggunakan dua metode yaitu *naive bayes classifier* dan *support vector machine*. Data dari permasalahan dapat diperoleh dari suatu jurnal, situs dan media elektronik ataupun buku elektronik.

3.3.3 Pengumpulan Data

Penelitian ini melakukan pengambilan data *tweet* dari *Twitter* terkait permasalahan kasus *bullying* yang dilakukan oleh Idol K-Pop Kim Garam dengan kata kunci ‘kim garam’, ‘bullying’, ‘kasus bullying’. Data yang digunakan diambil berdasarkan waktu trending dari awal mula rumor kasus bullying Kim Garam muncul yaitu pada tanggal 5 April 2022 Dalam pengumpulan data ini, peneliti mengambil data dengan cara crawling menggunakan *google colab*. *Crawling data* twitter merupakan proses pengambilan atau mengunduh data dari *server twitter* [22]. Hasil pengumpulan data akan disimpan dalam file bertipe *xlsx*.

3.3.4 Text Pre-processing

Setelah data diperoleh maka dilanjutkan ke tahap *text preprocessing* yang bertujuan untuk melakukan pembersihan suatu kata yang tidak diperlukan dan kata yang kurang bermakna untuk setiap data tweet. Tahapan *text preprocessing* yang digambarkan pada diagram Gambar 3.2.



Gambar 3. 2 Diagram alir tahap *text preprocessing*

a. *Cleaning*

Tahap *cleaning* dilakukan dengan tujuan untuk menghilangkan tanda baca, angka, URL, simbol, taggar, *username* atau *mention*, *emoticon*, *Retweet* dan karakter lain yang tidak diperlukan. Contoh hasil dari tahap *cleaning* terlihat di Tabel 3.1.

Tabel 3. 1 Contoh tahapan *Cleaning*

Data sebelum <i>cleaning</i>	Data setelah <i>cleaning</i>
Masih gak percaya kalo bener Kim garam pelaku. Kecewa gue.	Masih gak percaya kalo bener Kim garam pelaku Kecewa gue

b. *Case Folding*

Case folding merupakan tahapan untuk menggantikan semua huruf pada teks menjadi huruf kecil (*lowercase*) secara keseluruhan. Hasil *case folding* dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Contoh tahap *case folding*

Data sebelum <i>case folding</i>	Data setelah <i>case folding</i>
Masih gak percaya kalo bener Kim garam pelaku Kecewa gue	masih gak percaya kalo bener kim garam pelaku kecewa gue

c. *Tokenizing*

Tahap selanjutnya yaitu *tokenizing* bertujuan untuk menguraikan kalimat dalam teks menjadi sebuah kata per kata. Contoh hasil tahap tokenisasi ditunjukkan pada Tabel 3.3.

Tabel 3. 3 Contoh tahap *tokenizing*

Data sebelum <i>tokenizing</i>	Data setelah <i>tokenizing</i>
masih gak percaya kalo bener kim garam pelaku kecewa gue	‘masih’, ‘gak’, ‘percaya’, ‘kalo’, ‘bener’, ‘kim’, ‘garam’, ‘pelaku’, ‘kecewa’, ‘gue’

d. *Normalization*

Normalization perlu dilakukan untuk mengganti kata tidak standar atau yang tidak baku menjadi kata baku dan kata yang salah ejaan ataupun disingkat menjadi bentuk tertentu. Contoh hasil *normalization* seperti pada Tabel 3.4.

Tabel 3. 4 Contoh tahap *normalization*

Data sebelum <i>normalization</i>	Data setelah <i>normalization</i>
'masih', 'gak', 'percaya', 'kalo', 'bener', 'kim', 'garam', 'pelaku', 'kecewa', 'gue'	'masih', 'tidak', 'percaya', 'kalau', 'benar', 'kim', 'garam', 'pelaku', 'kecewa', 'saya'

e. *Stopword Removal*

Tujuan langkah ini adalah untuk menghilangkan kata-kata yang tidak memiliki arti atau arti penting, seperti dan, atau. Tabel 3.5 menunjukkan contoh hasil penerapan *stopword removal*.

Tabel 3. 5 Contoh Tahap *stopword removal*

Data sebelum <i>stopword</i> <i>removal</i>	Data setelah <i>stopword</i> <i>removal</i>
'masih', 'tidak', 'percaya', 'kalau', 'benar', 'kim', 'garam', 'pelaku', 'kecewa', 'saya'	'masih', 'tidak', 'percaya', 'benar', 'pelaku', 'kecewa'

f. *Stemming*

Tahap *stemming* bertujuan untuk memperbaiki sebuah kata kembali bentuk kata dasar dengan menghilangkan kata imbuhan depan dan belakang setiap kata. Contoh hasil implementasi *stemming* seperti pada Tabel 3.6.

Tabel 3. 6 Contoh tahap *stemming*

Data sebelum <i>stemming</i>	Data setelah <i>stemming</i>
'masih', 'tidak', 'percaya', 'benar', 'pelaku', 'kecewa'	'masih', 'tidak', 'percaya', 'benar', 'pelaku', 'kecewa'

3.3.5 Pelabelan Data

Dalam pengklasifikasian data masalah yang sering ditemukan yaitu ketidakseimbangan data [28]. Ketika suatu kelas memiliki jumlah jauh lebih besar dibanding dengan kelas lain dapat terjadi ketidakseimbangan data dan akan menyebabkan menurunnya kinerja klasifikasi pada kelas minoritas. Pada proses pelabelan data dibagi menjadi tiga kelas yaitu positif, negatif dan netral. Agar mempermudah pengklasifikasian, pelabelan untuk kelas positif dengan label 0, sedangkan untuk kelas negatif dengan label 1 dan kelas netral dengan label 2. Contoh hasil dari pelabelan data terlihat di Tabel 3.7.

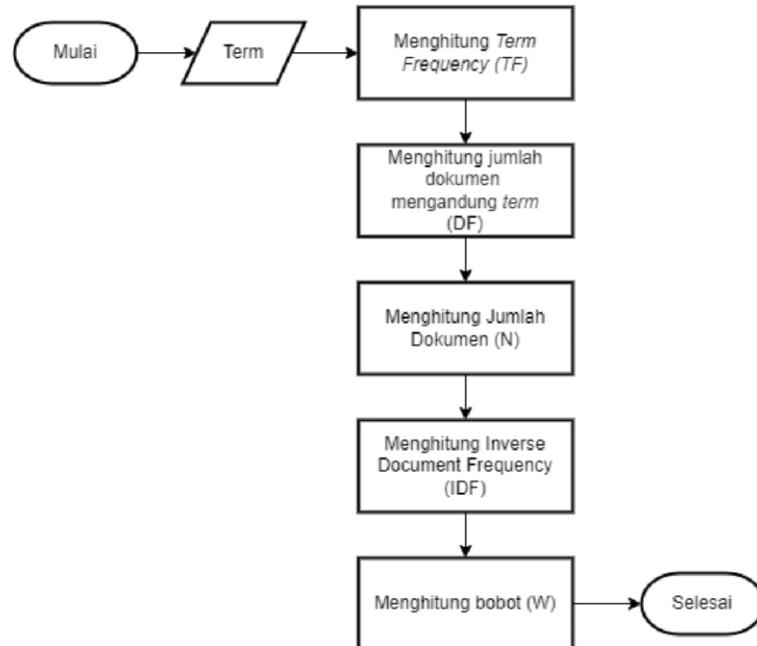
Tabel 3. 7 Pelabelan data

No.	Tweet	Label
1	Kecewa kalo kim garam beneran jadi pelaku bullying. Harusnya agency cari	1

No.	Tweet	Label
	tau masa lalu trainee dulu baru dimasukin ke line up debut.	
2	Agensinya benar-benar buruk, penggemarnya juga, mereka menunjukkan levelnya.	1
3	Kalau rumornya salah, kasihan banget asli kim garam gak salah apa apa tiba tiba dihujat. Mana dihujatnya sama netijen indo doang lagi, parah	0
4	Gw kepo, kalo kim garam beneran pelaku bullying apakah bakal tetap suka? kan suara sama dance nya dia ngga tau. Suka nya dari apa selain cantik...	2

3.3.6 Feature Selection TF-IDF

Data dari hasil *preprocessing* selanjutnya melalui proses *feature selection* dengan pembobotan TF-IDF. Tahap ini bertujuan untuk membantu meningkatkan nilai akurasi dalam klasifikasi. TF-IDF menghitung seberapa relevan sebuah kata dalam suatu dokumen yang ada. Pada dasarnya perhitungan bobot TF-IDF bekerja dengan menentukan frekuensi suatu kata dan membandingkannya dengan proporsinya di semua dokumen [24]. Proses perhitungan TF-IDF diilustrasikan seperti Gambar 3.3.



Gambar 3. 3 Tahapan Ekstraksi Fitur TF-IDF

Perhitungan bobot dilakukan dengan TF-IDF. Beberapa langkah dalam menentukan nilai bobot diantaranya adalah sebagai berikut:

- a. Menghitung *term frequency* (TF)

Menghitung berapa kali kemunculan sebuah *term* dalam dokumen yang dimaksud. TF dihitung dengan persamaan rumus (3.1)

$$tf_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}} \quad (3.1)$$

Keterangan :

$tf_{i,j}$: jumlah frekuensi *term*

$n_{i,j}$: jumlah *term* i dalam dokumen j

b. Menghitung jumlah dokumen mengandung term (DF)

Nilai DF diperoleh dengan menjumlahkan setiap kemunculan *term* yang sama pada seluruh dokumen.

c. Menghitung jumlah dokumen (N)

Dokumen akan dijumlahkan untuk mengetahui banyaknya keseluruhan dokumen yang ada. Jumlah dokumen direpresentasikan dengan “N”.

d. Menghitung *inverse document frequency* (IDF)

IDF adalah ukuran kemampuan *term* untuk membedakan kategori. Nilai IDF diberikan oleh persamaan (3.2)

$$idf_i = \log \frac{N}{df_i} \quad (3.2)$$

Keterangan :

idf_i : *Inverse Document Frequency*

N : Jumlah dokumen

df_i : Jumlah dokumen yang mengandung *term* (f_i)

e. Menghitung bobot (W)

Bobot TF-IDF adalah hasil perkalian bobot frekuensi term dengan invers frekuensi dokumen term. Nilai bobot dengan menggunakan TF-IDF dapat diperoleh dengan persamaan rumus (3.3) sebagai berikut:

$$W_{i,j} = tf_{i,j} \times idf_i \quad (3.3)$$

Keterangan :

$W_{i,j}$: Bobot TF-IDF

$tf_{i,j}$: Frekuensi *term*

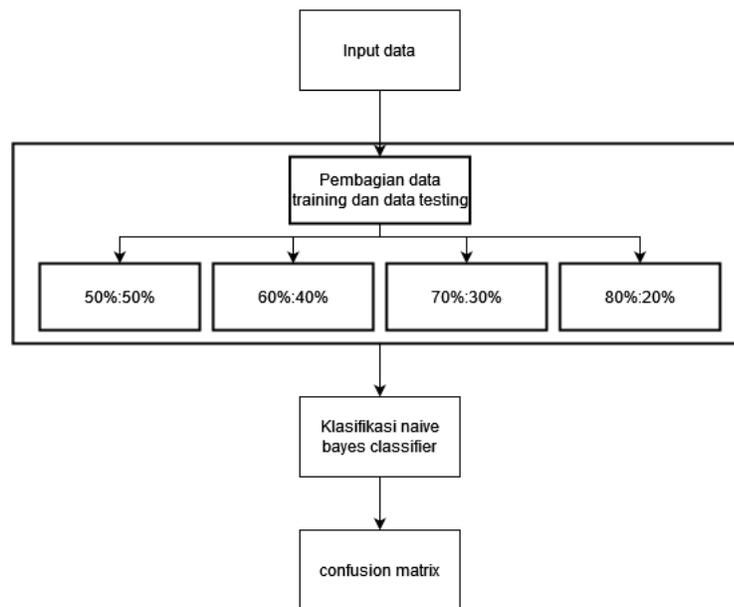
idf_i : *Inverse Document Frequency*

3.3.7 Klasifikasi model *naive bayes classifier* dan *support vector machine*

Tahap penelitian selanjutnya adalah membangun model klasifikasi dengan metode *naive bayes classifier* dan *support vector machine*. Implementasi pengujian dataset penelitian menggunakan bahasa pemrograman *python*. Pada model klasifikasi, dataset yang telah didapatkan akan dibagi menjadi data *training* dan *testing* dengan beberapa perbandingan.

1. *Naive bayes classifier*

Penerapan *naive bayes classifier* untuk klasifikasi data respon pengguna twitter terhadap kasus bullying oleh Kim Garam.

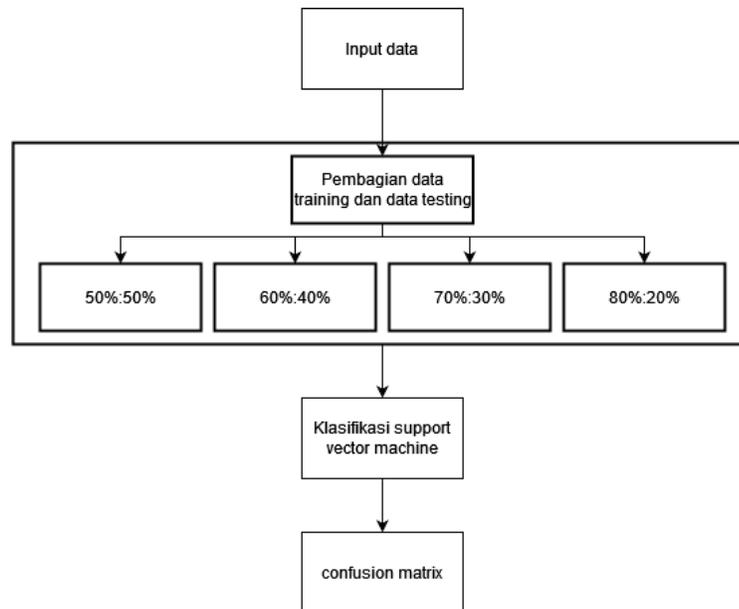


Gambar 3. 4 Alur *naive bayes classifier*

- a. Masukan data dan menentukan variabel atau fitur dalam penelitian. Fitur pada penelitian ini yaitu dokumen sampel tweet untuk dikelompokkan ke dalam dua kelas positif dan negatif.
- b. Dataset dibagi menjadi data *training* dan data *testing*. Penelitian ini menggunakan dua perbandingan berbeda yaitu data *training* sebanyak 50%, 60%, 70% dan 80%. Sedangkan data *testing* sebanyak 20%, 30%, 40% dan 50%.
- c. Membuat model *naive bayes classifier* yang akan digunakan untuk melakukan klasifikasi.
- d. Melihat *confusion matrix* untuk memudahkan dalam mengetahui apakah terdapat kesalahan dalam pengklasifikasian.

2. Support vector machine

Penerapan *support vector machine* untuk klasifikasi data respon pengguna twitter terhadap kasus *bullying* oleh Kim Garam.



Gambar 3. 5 Alur support vector machine

- a. Masukkan data dan menentukan variabel atau fitur dalam penelitian. Fitur pada penelitian ini yaitu dokumen sampel tweet untuk dikelompokkan ke dalam dua kelas positif dan negatif.
- b. Dataset dibagi menjadi data *training* dan data *testing*. Penelitian ini menggunakan dua perbandingan berbeda yaitu data *training* sebanyak 50%, 60%, 70% dan 80%. Sedangkan data *testing* sebanyak 20%, 30%, 40% dan 50%.
- c. Membuat model *support vector machine* yang akan digunakan untuk melakukan klasifikasi.

- d. Melihat *confusion matrix* untuk memudahkan dalam mengetahui apakah terdapat kesalahan dalam pengklasifikasian.

3.3.8 Evaluasi dan Analisis Performa Klasifikasi

Parameter yang digunakan pada tahap evaluasi yaitu *confusion matrix*. Evaluasi dilakukan untuk melihat nilai dari *accuracy*, *recall*, *precision* dan *f1-score* dengan melihat hasil performa dari masing-masing model klasifikasi *naive bayes classifier* dan *support vector machine* dari respon kasus *bullying* oleh Idol K-Pop Kim Garam. Setelah menerima hasil evaluasi, maka dapat menarik kesimpulan akhir dari penelitian.