

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Subjek dan Objek Penelitian**

Pada penelitian ini subjek penelitian yang akan diamati yaitu komentar masyarakat terhadap pelaksanaan KTT G20 pada media sosial Youtube. Objek permasalahan yang digunakan dalam penelitian ini adalah sudut pandang masyarakat mengenai pelaksanaan KTT G20 yang dilihat dari komentar konten video pada sosial media Youtube yang membahas KTT G20 yang diselenggarakan di Bali, Indonesia pada tanggal 15 November 2022.

#### **3.2 Alat dan Bahan Penelitian**

Pada penelitian ini membutuhkan alat dan bahan sebagai berikut: Alat:

1. Perangkat keras

Perangkat keras yang berupa laptop dengan spesifikasi:

- a. Acer Aspire 3*
- b. Processor AMD Ryzen 5*
- c. RAM 8GB, SSD 128 GB + HDD 1000 GB*
- d. AMD Radeon Vega Mobile Graphics*
- e. Operating Sytems Windows 10*

2. Perangkat lunak

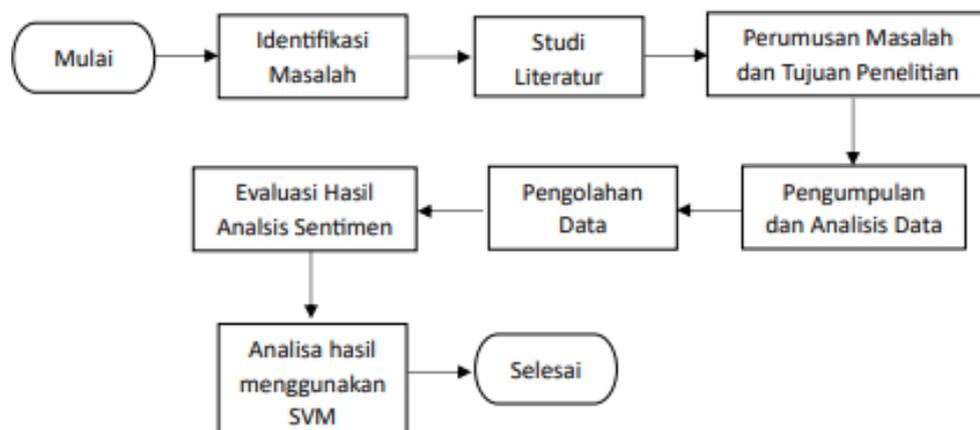
Perangkat lunak yang digunakan sebagai berikut:

- a. Google Colab*
- b. Browser (Google Chrome)*
- c. Microsoft word 2021*
- d. Python*

3. Bahan:
  - a. Data dari komentar sosial media Youtube.
  - b. Jaringan internet

### 3.3 Diagram Alir Penelitian

Tahapan penelitian pertama dilakukan menentukan permasalahan dari topik yang diambil. Lalu menentukan tujuan dilakukannya penelitian dari topik yang ditemukan. Selanjutnya melakukan studi literatur untuk mempelajari lebih dalam mengenai topik yang akan diangkat dalam penelitian. Setelah itu melakukan pengumpulan data yang akan digunakan dalam proses penelitian yang didapat dari komentar pada sosial media Youtube. Dari dataset yang diperoleh akan dianalisis untuk mengetahui isi dari data setelah dataset didapatkan. Selanjutnya pengolahan dataset mentah yang diolah menjadi data yang siap untuk digunakan penulis. Setelah melakukan olah data maka data tersebut dianalisis untuk menemukan sentimen menggunakan metode *Support Vector Machine* (SVM) dan mencari nilai akurasi dari 4 kernel algoritma SVM. Terakhir hasil analisis sentimen tersebut akan ditarik sebuah kesimpulan yang berupa evaluasi selama penelitian berjalan dan hasil yang dapat digunakan peneliti selanjutnya untuk pengembangan teori yang sudah diteliti pada penelitian selanjutnya.



**Gambar 3. 1 Proses Penelitian**

Pada Gambar 3.1 merupakan flowchart *terminator symbol* yang berbentuk oval terletak di awal dan akhir process dimana symbol tersebut untuk permulaan (start) atau akhir (stop) dari suatu kegiatan. Sedangkan *processing symbol* (persegi panjang) yaitu simbol yang menunjukkan pengolahan yang dilakukan oleh komputer.

### **3.3.1 Identifikasi Masalah**

Pada tahap awal melakukan identifikasi terhadap masalah yang ada di sekitar lingkungan. Dimana masalah tersebut akan dijadikan topik yang akan dikaji dalam penelitian. Dari masalah yang akan dikaji tersebut dilakukan pencarian solusi yang tepat. Masalah yang diidentifikasi dalam penelitian yang akan dilakukan adalah opini masyarakat terhadap KTT G20 yang dilaksanakan di Bali, Indonesia pada komentar sosial media Youtube. Oleh karena dalam komentar video pada Youtube pasti terdapat opini masyarakat yang positif dan negatif maka dilakukan analisis sentimen berdasarkan opini tersebut.

### **3.3.2 Studi Literatur**

Studi literatur memiliki tujuan untuk mencari referensi sebagai pendukung untuk memperoleh informasi, kajian dan teori dari penelitian terdahulu dalam landasan penelitian yang akan dilakukan bersumber dari jurnal, internet, buku dan sebagainya. Studi literatur yang dikaji memiliki keterkaitan dengan penelitian yang akan dibahas dari segi metode ataupun dari topik penelitiannya. Studi literatur yang digunakan dalam penelitian memiliki rentang waktu dari 5 tahun kebelakang yaitu sekitar tahun 2018-2023.

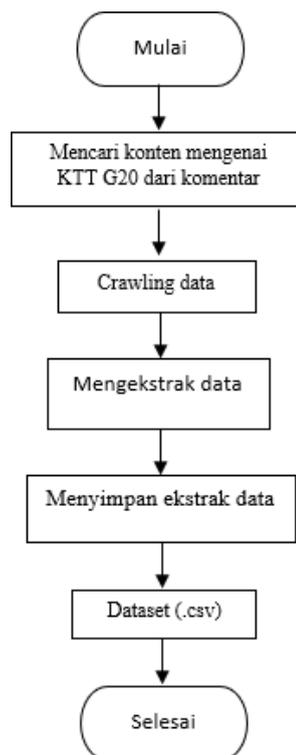
### **3.3.3 Perumusan Masalah dan Tujuan Penelitian**

Perumusan masalah didapatkan setelah melakukan observasi atau identifikasi masalah terhadap objek yang akan dibahas dan kemudian dari hasil permasalahan tersebut dijadikan topik penelitian. Rumusan masalah yang didapat dijadikan batasan dalam penelitian agar penelitian memiliki fokus yang jelas terhadap topik yang sudah ditetapkan. Selanjutnya adalah menetapkan metode dalam penelitian yang dilakukan menggunakan algoritma SVM untuk menyelesaikan analisis

sentimen dari topik yang diambil. Tujuan penelitian ditetapkan agar penelitian memiliki arah yang jelas dalam pelaksanaannya

### 3.3.4 Pengumpulan dan Analisis Data

Data yang digunakan dalam penelitian ini adalah data komentar dari sosial media Youtube mengenai KTT G20 pada tanggal 15 November 2022. Tahap dalam analisis sentimen adalah yang pertama mencari konten yang sesuai topik pada Youtube di dalam penelitian ini konten Youtube yang digunakan yaitu video pada channel Sekretariat Presiden mengenai pelaksanaan KTT G20. Kemudian konten video tersebut dilakukan *crawling data* menggunakan Youtube API yang dibuat menggunakan *api\_keys* yang di dapat pada *google cloud console* untuk selanjutnya diproses dengan bantuan *google colab* setelah mendapatkan *api\_key* dan link video yang dibutuhkan kemudian dilakukan ekstrak data dan disimpan dalam format CSV.



**Gambar 3. 2 Diagram alir pengumpulan data**

Data yang sudah disimpan dalam format CSV setelah proses *crawling* kemudian di *upload* pada *Google colab* untuk kemudian dapat digunakan dalam proses *pre-processing* pada tahap pengolahan data agar lebih mudah untuk dianalisis dan menghasilkan fitur yang optimal dari data agar mudah diproses.

	publishedAt	authorDisplayName	textDisplay	likeCount
0	2023-08-01T07:11:26Z	Kuningan Clawigebang	Kalo indonesia jadi tuan rumah piala dunia ker...	0
1	2023-07-31T11:17:12Z	Zahran S.f	Keren	0
2	2023-07-31T10:39:05Z	Jun,"Tanpa Jokowi akan te...	NaN	NaN
3	2023-07-31T08:48:36Z	Runi Clear	memang luar biasa acara yang arahin pak wishut...	0
4	2023-07-30T20:13:30Z	cantra natiphi	g20 tapi kok warganya ngetops jadi kuli di ran...	0
...	...	...	...	...
7091	2022-11-15T12:59:11Z	jrambah bang	🤔	0
7092	2022-11-15T12:55:24Z	Zero Death	first	1
7093	2022-11-15T12:55:24Z	Yanti Gani	Sukses	7
7094	2022-11-15T12:55:20Z	IBNU JiteX	Yahhh dh lesai	2
7095	2022-11-15T12:55:16Z	RICKO DAVID	music video	7

7096 rows x 4 columns

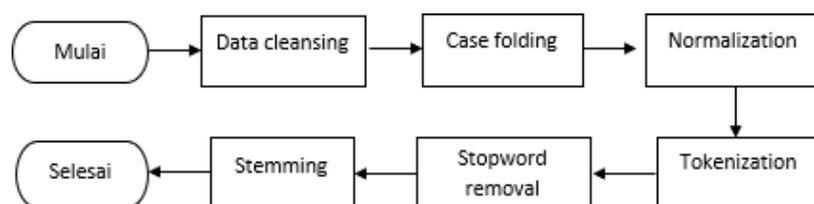
**Gambar 3.3** Total dataset yang dikumpulkan

Setelah mendapatkan informasi melalui pengumpulan data yang ditampilkan pada Gambar 3.3 menunjukkan jumlah dataset yang diperoleh dari *crawling* data sebelum tahap *pre-processing*. Dataset yang dikumpulkan sejumlah 7096 data. Data yang diperoleh dan diolah tersebut akan dijadikan data latih dan data uji untuk menentukan label sentimen dengan 3 skenario yaitu 70:30, 75:25, 80:20. Dalam analisis data yang dilakukan ditemukan adanya data yang bernilai kosong atau NaN, karakter, tanda baca dan emoji.

### 3.3.5 Pengolahan Data

Dataset yang telah didapatkan dari proses pengumpulan data dan dilakukan analisis terhadap dataset terhadap komentar masyarakat pada media sosial Youtube selanjutnya diolah untuk dilakukan tahap *pre-processing* pada Gambar 3.4 dan pembobotan TF-IDF.

#### 1. Pre-Processing



**Gambar 3.4** Tahap *pre-processing*

Tahap ini data yang didapat akan dilakukan pembersihan data yang tidak digunakan (*data cleansing*), *case folding* untuk menyamaratakan teks dalam menggunakan huruf kapital yang akan diubah keseluruhan menjadi huruf kecil[35], melakukan normalisasi data supaya sesuai dengan KBBI[36], tokenisasi dengan membagi teks menjadi bagian atau token tertentu[35], *stopwords removal* dimana kata hasil token dipilih untuk mengambil kata-kata yang penting yaitu kata yang akan digunakan untuk mewakili dokumen, *stemming* yaitu proses untuk menghasilkan kata dasar sehingga lebih relevan menggunakan *library* sastrawi untuk melakukan *stemming* berbahasa Indonesia[37], dan menampilkan ikhtisar untuk menyaring teks berupa kata-kata yang memiliki frekuensi yang tinggi. Data yang didapat dilakukan pelabelan untuk membagi dua kelas data menjadi sentimen positif dan negatif menggunakan data latih dan data uji dengan pembagian data menjadi tiga skenario yaitu 80:20, 70:30, dan 75:25 pada 4 masing-masing kernel SVM yaitu *linear*, *sigmoid*, *polynomial*, dan *Rbf*. Data latih digunakan untuk pembelajaran sistem dan dijadikan acuan. Sedangkan data uji untuk mengetahui letak kelas suatu data[38].

## 2. Pembobotan TF-IDF

Tahap ini akan menyeleksi fitur menggunakan pembobotan TF-IDF untuk mengetahui banyaknya data yang muncul pada dokumen dan dapat meningkatkan akurasi pada klasifikasi. Dalam proses pembobotan TF-IDF menggunakan *library sklearn*. Pembobotan fitur merupakan suatu cara atau metode untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen[37]. Dimana TF merupakan frekuensi munculnya kata (*term*) dalam dokumen, dan DF adalah frekuensi munculnya kata (*term*) dari jumlah dokumen[39].

### 3.3.6 Analisa Hasil Algoritma *Support Vector Machine*

Tahap ini merupakan tahap klasifikasi sentimen menggunakan algoritma *Support Vector Machine (SVM)*. SVM akan memetakan variabel data menjadi

beberapa kelas menggunakan *hyperplane* yang optimal dengan memaksimalkan jarak tiap kelas[40].

**Tabel 3.1 Contoh Perhitungan SVM**

$x_1$	$x_2$	Kelas (y)	Support Vector (SV)
1	1	1	1
1	-1	-1	1
-1	1	-1	1
-1	-1	-1	0

Pada tabel diatas memiliki 4 data yang setiap data memiliki nilai  $x_1$ ,  $x_2$ , kelas (y), dan *Support Vector* (SV). Terdapat dua fitur yaitu  $x_1$  dan  $x_2$  sehingga  $w$  juga memiliki 2 fitur yaitu  $w_1$  dan  $w_2$ . Teknik klasifikasi fitur yang disebut SVM berupaya mengidentifikasi *hyperplane* yang ideal. Untuk menentukan *hyperplane* tersebut yaitu menggunakan rumus berikut: Meminimalkan nilai

$$\frac{1}{2} ||w||^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (3.1)$$

Syarat:

$$y_1(w \cdot x_1) \geq 1, \quad i = 1, 2, 3, \dots, N$$

$$y_1(w_1 \cdot x_1 + w_2 \cdot x_2 + b) \geq 1$$

Contoh perhitungan:

$$y_1(w_1 \cdot x_1 + w_2 \cdot x_2 + b) \geq 1$$

$$1(w_1 \cdot 1 + b) \geq 1$$

$$(w_2 \cdot x_2 + b) \geq 1$$

Hasil:

$$1. (w_1 + w_2 + b) \geq 1, \text{ untuk } y_1 = 1, x_1 = 1, x_2 = 1$$

$$2. (-w_1 + w_2 - b) \geq 1, \text{ untuk } y_2 = -1, x_1 = 1, x_2 = -1$$

$$3. (w_1 + w_2 - b) \geq 1, \text{ untuk } y_3 = -1, x_1 = -1, x_2 = 1$$

$$4. (w_1 + w_2 - b) \geq 1, \text{ untuk } y_4 = -1, x_1 = -1, x_2 = -1$$

Kemudian jumlahkan persamaan sebagai berikut:

3. Persamaan (1) dan (2)

$$(w_1 + w_2 + b) \geq 1 \quad (-w_1 + w_2 - b) \geq 1$$

$$\hline +$$

$$2w_2 = 2$$

$$\text{Maka } w_2 = 1$$

4. Persamaan (1) dan (3)

$$(w_1 + w_2 + b)$$

$$\geq 1 \quad (w_1 - w_2$$

$$- b) \geq 1$$

$$\hline +$$

$$2w_1 = 2$$

$$\text{Maka } w_1 = 1$$

5. Persamaan

$$(2) \text{ dan } (3) \quad (-w_1 + w_2 - b) \geq 1 \quad (w_1 - w_2 - b) \geq 1$$

$$\hline +$$

$$-2b = 2$$

$$\text{Maka } b = -1$$

Sehingga persamaan hyperplanenya menjadi

$$w_1x_1 + w_2x_2 + b = 0$$

$$w_1 + x_1 - 1 = 0$$

$$x_2 = 1 - x_1$$

### 3.3.7 Evaluasi Hasil Analisis Sentimen

Evaluasi yang dilakukan adalah menguji bagaimana algoritma yang digunakan dalam penelitian yaitu algoritma SVM dapat melakukan performa model dengan baik atau tidak. Data latih yang sudah diterapkan dengan 3 skenario uji yaitu 70:30, 75:25 dan 80:20 untuk melihat keakuratan dari klasifikasi. Kemudian membandingkan dua dataset yaitu antara data latih dan data uji dengan pengaplikasian empat kernel terbaik SVM yaitu *linier*, *polynomial*, RBF, dan *sigmoid* sehingga dapat menghasilkan nilai *accuracy*, *precission*, *recall*, dan *F1-Score* dengan menggunakan metode *confussion matrix*[38].