

BAB III

METODOLOGI PENELITIAN

3.1 Objek dan Subjek Penelitian

3.1.1 Subjek Penelitian

Subjek pada penelitian ini merupakan pengguna twitter yang melakukan tweet mengenai tragedi kanjuruhan dan juga penonton youtube yang berkomentar pada konten berjudul Tragedi Kanjuruhan #UsutsampaiTuntas di channel Mata Najwa.

3.1.2 Objek Penelitian

Objek penelitian ini adalah *tweet* pengguna twitter terhadap tragedi kanjuruhan mulai dari 1 Oktober – 31 Oktober 2022 dan komentar penonton youtube pada konten yang berjudul Tragedi Kanjuruhan #UsutsampaiTuntas di *channel* Mata Najwa mulai dari awal tayang sampai 20 November 2022.

3.2 Alat dan Bahan Penelitian

3.2.1 Alat Penelitian

Dalam melakukan penelitian ini peneliti menggunakan perangkat keras yaitu laptop dengan sistem operasi dan spesifikasi hardware seperti berikut:

Tabel 3. 1 Alat Penelitian

Merek	Hp Notebook
Sistem Operasi	Ubuntu 22.04 dan Windows 10 Pro 64-bit
<i>Processor</i>	Intel(R) Core (TM) i5-6200U (4 CPUs) 2,4 GHz + AMD Radeon (TM) R5 M330
RAM	12 GB
SSD	256

Penelitian ini juga menggunakan beberapa perangkat lunak seperti:

1. *Ubuntu 20.04*
2. *Windows 11*
3. *Jupyter notebook*
4. *Google Collab*
5. *Python 3*
6. *Twitter*
7. *Youtube*
8. *Browser Chrome*

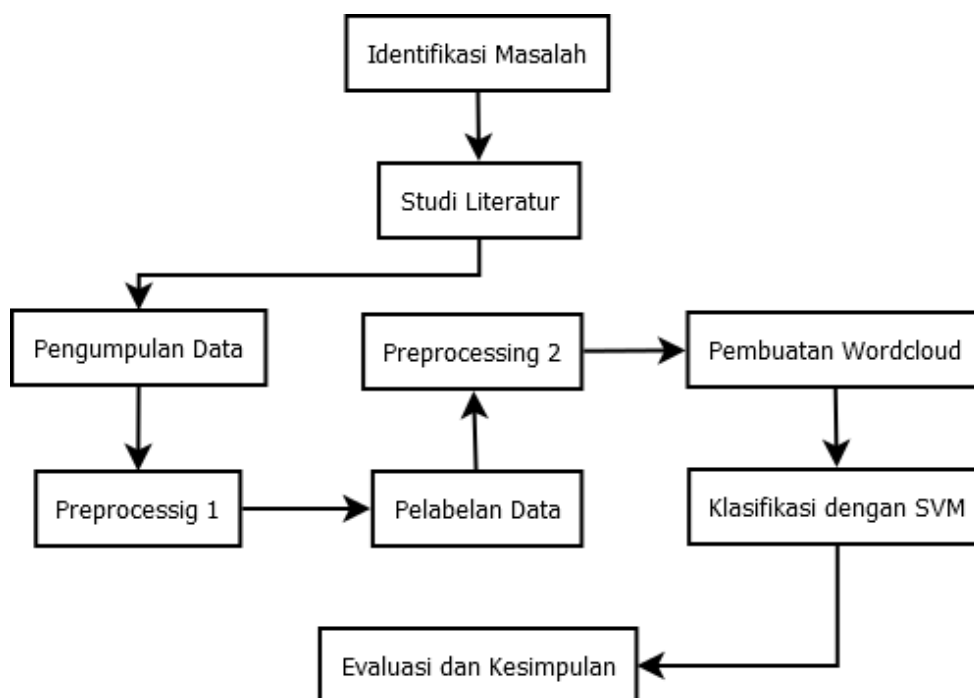
3.2.2 Bahan Penelitian

Bahan penelitian yang digunakan pada penelitian ini merupakan data-data sentimen masyarakat terhadap tragedi kanjuruhan yang meliputi data sebagai berikut:

1. Data penelitian diambil dari data tweet pengguna *twitter* mengenai tragedi kanjuruhan dan data komentar *youtube* “Tragedi Kanjuruhan #UsutsampaiTuntas | Mata Najwa”.
2. *Tweet* yang diambil adalah *tweet* dari 1 oktober - 31 oktober 2022 dan komentar *youtube* “Tragedi Kanjuruhan #Usutsampai Tuntas | Mata Najwa” dari 6 Oktober sampai pada tanggal 20 November 2022.

3.3 Diagram Alir Penelittian

Penelitian ini dilakukan dengan melewati beberapa tahap yang dimulai dari identifikasi masalah, menyusun latar belakang masalah, *crawling* data, *preprocessing* data, pembuatan *wordcloud*, pemodelan dan pembobotan sentimen, klasifikasi dengan menggunakan *SVM*, melihat hasil dan evaluasi kemudian menarik kesimpulan dan saran. Diagram alur dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Diagram Alir Penelitian

3.3.1 Identifikasi dan Perumusan masalah

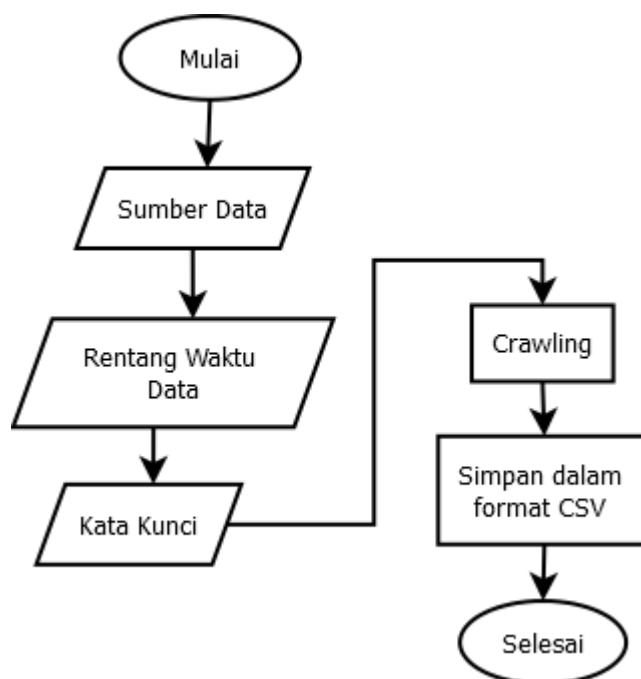
Proses Mengidentifikasi dan mendefinisikan masalah yang akan ditangani dalam penelitian merupakan proses yang menggunakan pendekatan logika dan matematika. Dengan demikian, tujuan dari proses ini adalah untuk menyusun pertanyaan mendasar yang tidak membatasi diskusi dalam penelitian. Ini penting untuk menghindari kesulitan dalam menemukan solusi yang tepat untuk masalah yang telah ditentukan.

3.3.2 Studi Literatur

Dalam penelitian ini, referensi atau sumber data merupakan dasar dan pedoman yang diperlukan untuk mengembangkan penelitian. Oleh karena itu, penulis melakukan banyak bacaan, penelitian, dan pemahaman terhadap konsep dan pembahasan terkait sentimen analisis dan algoritma *SVM* di jurnal-jurnal seperti *scopus*, *scholar*, dan sumber terpercaya lainnya. Hasil dari proses ini akan dijadikan sebagai dasar untuk menulis dan melakukan penelitian selanjutnya.

3.3.3 Pengumpulan Data

Dalam melakukan pengumpulan data terkait dengan tragedi kanjuruhan dilakukan proses seperti diagram alir dibawah ini.



Gambar 3. 2 Diagram Alir Pengumpulan Data

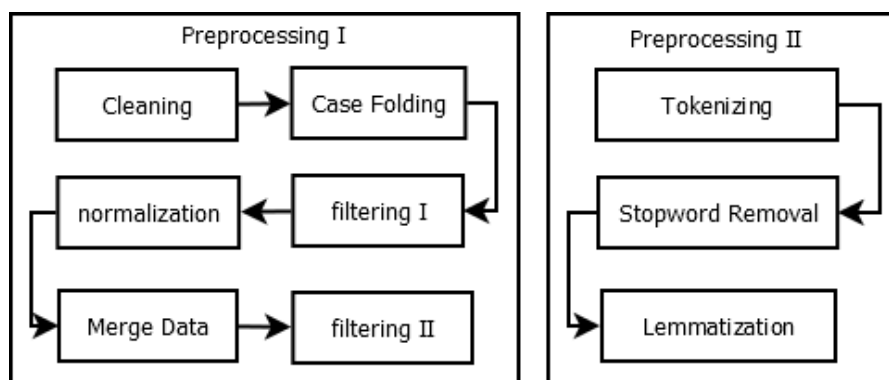
Pada Gambar 3.2 menjelaskan bagaimana tahapan pengumpulan data dilakukan. Data diambil dari dua sumber yaitu *Twitter* dan *Youtube*, data dari *Twitter* berupa *tweet* tentang tragedi kanjuruhan pada 1-31 Oktober 2022, sedangkan data dari *Youtube* berupa komentar pada konten *youtube* Tragedi Kanjuruhan #UsutsampaiTuntas | Mata Najwa dari awal unggah sampai 20 November 2022. *Crawling tweet* menggunakan *library snsrape*, sedangkan *crawling* komentar *Youtube* menggunakan *spreadsheet Apps Script* dengan *API Youtube*. Setelah selesai melakukan *crawling* maka data disimpan dalam format *CSV*. Hasil *crawling* yang didapat adalah 15.224 *tweet* dan 3999 komen *youtube*.

Tabel 3. 2 Contoh *Tweet*/komentar tentang tragedi kanjuruhan

No	Tweet/Komentar	Sumber
1	Indonesia berduka di Kanjuruhan. Ikut sedih untuk adanya puluhan korban jiwa, baik dari supporter maupun polisi dalam pertandingan Persebaya - Arema. Nyawa manusia terlalu berharga untuk dikorbankan hanya demi sepak bola. #kanjuruhandisaster #kanjuruhan https://t.co/nH9kC37HvC	Twitter #kanjuruhan
2	Selidiki selidiki, bener di selidiki yaaaaa pak 😊 ini nyawa manusia ratusan 🙄 melihat dari video ² yg ada sudah terbukti pak.. Tinggal di usut tuntas saja yg jelas siapa yang kunci pintu keluar, siapa yang menembakkan gas air mata sebanyak itu kemana mana, siapa yang memulai masuk ke lapangan. Dan TANYA KEJELASAN NYA!	Youtube Tragedi Kanjuruhan #UsutsampaiTuntas Mata Najwa

3.3.4 Preprocessing

Pada tahap *preprocessing* data di penelitian ini digambarkan seperti diagram alir pada Gambar 3.3 dibawah:

Gambar 3. 3 Diagram Alir *Preprocessing*

Pada Gambar 3.3 dijelaskan tahapan *preprocessing* data. *Preprocessing* terdiri dari dua bagian, yaitu *preprocessing* pertama sebelum pelabelan data dan *preprocessing* kedua setelah pelabelan data. Pada *preprocessing* pertama, dilakukan beberapa langkah. Pertama, dilakukan

cleaning untuk menghilangkan karakter-karakter yang tidak diperlukan seperti *tag*, *hashtag*, *link*, *emoji*, dan angka. Selanjutnya, dilakukan *case folding* untuk mengubah semua teks menjadi huruf kecil. Setelah itu, ada *filtering* I yang berfungsi menyaring tweet atau komentar yang tidak relevan dengan tragedi Kanjuruhan. Selanjutnya, dilakukan *normalization* untuk menormalisasi kata-kata, termasuk singkatan-singkatan, menjadi kata-kata sebenarnya. Setelah itu, dilakukan *Merge Data* untuk menggabungkan data hasil normalisasi dari *Twitter* dan *YouTube*. Penggabungan data dilakukan setelah *filtering* I dan normalisasi karena data dari *twitter* dan *youtube* memiliki *noise* yang beda sehingga dilakukan *filtering* I menggunakan keyword yang berbeda pula kemudian kedua data digabungkan. Terakhir, dilakukan *filtering* II untuk menyaring teks yang relevan dengan tragedi Kanjuruhan.

Pada *preprocessing* kedua setelah pelabelan data, langkah-langkahnya sebagai berikut. Pertama, dilakukan *tokenizing* untuk mengubah setiap kalimat menjadi token-token atau dipisahkan menjadi kata per kata. Selanjutnya, dilakukan *stopword removal* untuk menghilangkan kata-kata yang tidak penting atau tidak memiliki makna. Terakhir, dilakukan lemmatisasi untuk mengubah setiap kata menjadi bentuk dasarnya berdasarkan konteks dan makna yang dimaksud.

Tabel 3. 3 Contoh hasil *preprocessing* data

No	<i>Cleaned text</i>	<i>Tokenizing</i>	<i>Lematizing</i>
1	indonesia berduka di kanjuruhan ikut sedih untuk adanya puluhan korban jiwa baik dari supporter maupun polisi dalam pertandingan persebaya arema nyawa manusia terlalu berharga untuk dikorbankan hanya demi sepak bola	['indonesia', 'berduka', 'di', 'kanjuruhan', 'ikut', 'sedih', 'untuk', 'adanya', 'puluhan', 'korban', 'jiwa', 'baik', 'dari', 'supporter', 'maupun', 'polisi', 'dalam', 'pertandingan', 'persebaya', 'arema', 'nyawa', 'manusia', 'terlalu', 'berharga', 'untuk', 'dikorbankan', 'hanya', 'demi', 'sepak', 'bola']	indonesia berduka di kanjuruhan ikut sedih atas adanya puluhan korban jiwa baik dari supporter maupun polisi dalam pertandingan persebaya arema. nyawa manusia terlalu berharga untuk dikorbankan hanya demi sepak bola.

No	<i>Cleaned text</i>	<i>Tokenizing</i>	<i>Lematizing</i>
		'bola']	
2	selidiki selidiki bener di selidiki yaaaa pak ini nyawa manusia ratusan melihat dari video yg ada sudah terbukti pak tinggal di usut tuntas saja yg jelas siapa yang kunci pintu keluar siapa yang menembakkan gas air mata sebanyak itu kemana mana siapa yang memulai masuk ke lapangan dan tanya kejelasan nya	['selidiki', 'selidiki', 'bener', 'di', 'selidiki', 'yaaaa', 'pak', 'ini', 'nyawa', 'manusia', 'ratusan', 'melihat', 'dari', 'video', 'yg', 'ada', 'sudah', 'terbukti', 'pak', 'tinggal', 'di', 'usut', 'tuntas', 'saja', 'yg', 'jelas', 'siapa', 'yang', 'kunci', 'pintu', 'keluar', 'siapa', 'yang', 'menembakkan', 'gas', 'air', 'mata', 'sebanyak', 'itu', 'kemana', 'mana', 'siapa', 'yang', 'memulai', 'masuk', 'ke', 'lapangan', 'dan', 'tanya', 'kejelasan', 'nya']	selidiki selidiki benar di selidiki ya pak ini nyawa manusia ratusan melihat dari video yang ada sudah terbukti pak tinggal diusut tuntas saja yang jelas siapa yang kunci pintu keluar siapa yang menembakkan gas air mata sebanyak itu ke mana-mana siapa yang mulai masuk ke lapangan dan tanya kejelasannya

Tabel 3.3 kolom kedua adalah komentar yang sudah melewati proses *case folding*, dan penghilangan karakter yang tidak perlu, kemudian dilakukan *tokenizing* untuk memisahkan tiap kata pada kalimat, dan pada kolom keempat dilakukan lematisasi untuk mengubah kata ke bentuk dasarnya sesuai dengan konteks.

3.3.5 Pembuatan *Wordcloud*

Pembuatan *wordcloud* dilakukan agar dapat mengetahui frekuensi kemunculan suatu kata dalam dataset tragedi kanjuruhan. Jika kata dalam suatu dataset tersebut memiliki jumlah kemunculan yang tinggi maka pada *wordcloud* kata tersebut akan semakin membesar dan sebaliknya jika suatu kata dalam data set tersebut jarang muncul maka pada *wordcloud* kata tersebut akan semakin kecil atau bahkan tidak ditampilkan. Pembuatan *wordcloud* menggunakan *library python wordcloud* kemudian ditampilkan menggunakan *library pyhon matplotlib*.



Gambar 3. 4 *Wordcloud* dari *tweet* tragedi kanjuruhan

Gambar 3.4 menunjukkan visualisasi *wordcloud* dari *tweet* tragedi kanjuruhan, kata yang dominan ditunjukkan dengan huruf yang lebih besar seperti tragedi kanjuruhan, air mata, gas, polisi, dan diikuti kata-kata lainnya.

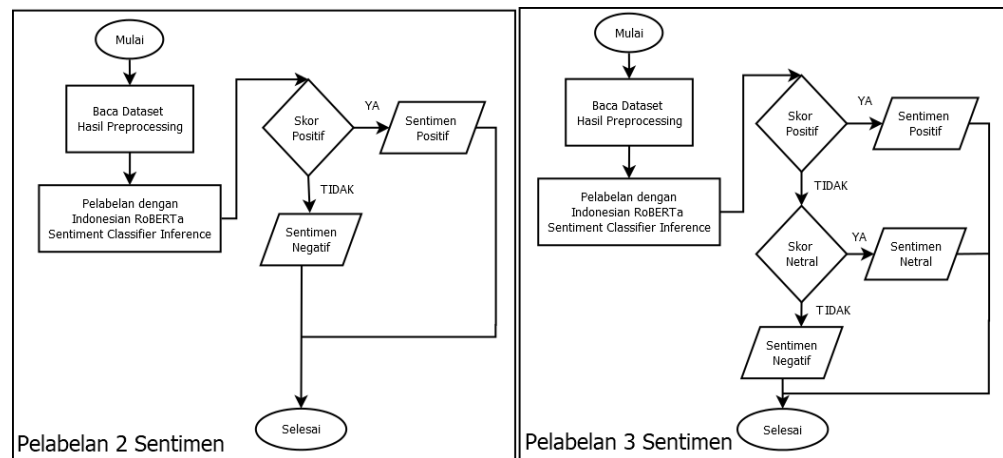


Gambar 3. 5 *Wordcloud* dari komen *youtube* Tragedi
Kanjuruhan|UsutsampaiTuntas

Gambar 3.5 menunjukkan visualisasi *wordcloud* dari komentar *youtube* Tragedi *Kanjuruhan|UsutsampaiTuntas*, kata yang dominan ditunjukkan dengan huruf yang lebih besar seperti gas air, air mata, aremania, supporter, polisi dan diikuti kata-kata lainnya.

3.3.6 Pelabelan Data

Pada tahap pelabelan data di penelitian ini digambarkan seperti diagram alir pada Gambar 3.6 dibawah:



Gambar 3. 6 Diagram Alir Pelabelan Data

Pada Gambar 3.6 menjelaskan bahwa data hasil *preprocessing I* selanjutnya akan dilabeli, pada penelitian ini menggunakan model *Indonesian RoBERTa Sentiment Classifier Inference* yang memiliki akurasi sebesar 95,36% dalam mengklasifikasi teks, komentar, ataupun *review* dalam bahasa Indonesia [26]. Model ini juga dipakai dalam penelitian mengenai analisis sentimen masyarakat mengenai vaksin Covid-19 [31]. Pelabelan akan menggunakan dua skenario yaitu pelabelan dengan dua sentimen dan pelabelan dengan 3 sentimen. Jika teks yang masuk pada model ini mendapat skor positif maka sentimen itu akan dilabeli sebagai sentimen positif, Jika sentimen yang masuk pada model ini mendapat skor netral maka sentimen itu akan dilabeli sebagai sentimen netral, dan jika bukan dari dua itu maka sentimen tersebut akan dilabeli sebagai sentimen negatif.

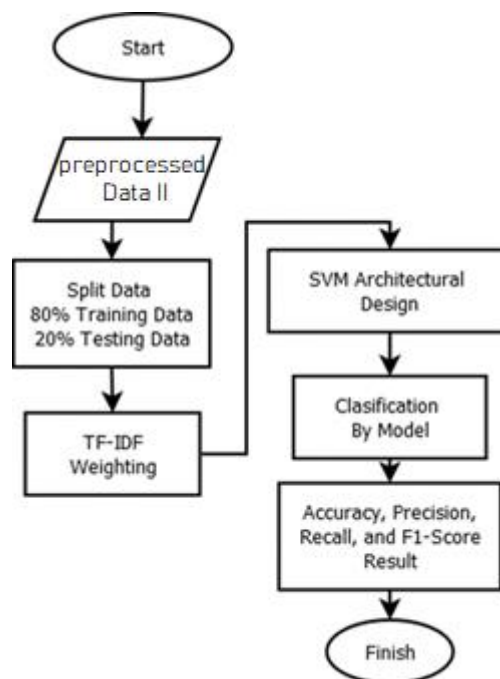
Tabel 3. 4 Contoh *dataset* yang sudah dilabeli

No	Text	Label
1	penggunaan gas air mata sesuai prosedur nyatanya banyak korban jadi bertanggung jawab salah sop benerin sop bukan malah jadi dasar pembenaran	Negatif
2	aku tau gak saling kenal kamu baca reply aku pengen bilang terima kasih banyak teman2 army bts berdonasi duka semua korban niat baik amp ketulusan kalian semoga dibalas	Positif
3	turut berdukacita sedalam dalam semua korban stadion kanjuruhan semoga kejadian tidak terulang	Netral

Tabel 3.4 merupakan contoh teks yang sudah dilabeli menggunakan model *deep learning*, mulai dari sentimen positif, negatif, ataupun netral.

3.3.7 Klasifikasi dengan SVM

Pada tahap pelabelan data di penelitian ini digambarkan seperti diagram alir pada Gambar 3.7 dibawah:



Gambar 3. 7 Diagram Alir Klasifikasi

Pada Gambar 3.7 menjelaskan proses klasifikasi dengan *SVM*, dataset yang sudah selesai dilabeli akan dibagi menjadi data latih 80% dan data uji 20%, kemudian dilakukan pembobotan dengan *TF-IDF*. Setelah itu dilakukan perancangan arsitektur *SVM*. Perancangan arsitektur *SVM* meliputi pengubahan jenis kernel *SVM* yang digunakan dan pengubahan nilai parameter dalam *SVM* dengan menggunakan *Grid Search* untuk mencari nilai parameter terbaik untuk setiap kernelnya. Setelah model selesai dirancang maka dilakukan klasifikasi dengan setiap model yang sudah dirancang. Hasil klasifikasi kemudian dievaluasi dengan *confusion matrix* yang meliputi nilai akurasi, presisi, *recall*, dan *f1-score*.

Tabel 3. 5 Contoh pembobotan dengan *TF-IDF*

Text	TF-IDF
penggunaan gas air mata sesuai prosedur nyatanya banyak korban jadi bertanggung jawab salah sop benerin sop bukan malah jadi dasar pembenaran	(0,11621) 0.35260290338351147
	(0,2051) 0.35260290338351147
	(0, 11947) 0.2255692196407906
	(0, 3469) 0.2435615270596058
	(0, 13500) 0.21920032957660687
	(0, 12863) 0.15136150807519105
	(0, 2676) 0.15512127157112215
	(0, 6690) 0.17359688852511582
	(0, 14588) 0.1745941023691693
	(0, 8788) 0.19531507122189276
	(0, 2029) 0.19454513169912888
	(0, 9038) 0.12791460778498873
	(0, 824) 0.12840647560207805
	(0, 5100) 0.12672806552442517
	(0, 6575) 0.21732578350410445
	(0, 10544) 0.1928681566326844
	(0, 1800) 0.15472134845681404
(0, 7904) 0.10351547402531716	
(0, 13977) 0.4914357395698443	

Tabel 3.5 menunjukkan contoh kalimat yang diubah kedalam bentuk vektor menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* sesuai dengan rumus (2.1). Proses ini bertujuan untuk mengubah setiap kata dalam kalimat menjadi vektor numerik dengan memanfaatkan nilai frekuensi kemunculan kata dalam kalimat (*TF*) dan nilai kebalikan dari frekuensi kemunculan kata tersebut dalam seluruh dokumen (*IDF*).

Tabel 3. 6 Parameter kernel *SVM* yang digunakan

Kernel	Parameter			
	<i>C</i>	<i>Gamma</i>	<i>Degree</i>	<i>Coef</i>
<i>Linear</i>	✓			
<i>RBF</i>	✓	✓		
<i>Polynomial</i>	✓	✓	✓	✓
<i>Sigmoid</i>	✓	✓		✓

Tabel 3.6 menunjukkan parameter yang digunakan untuk menguji fungsi kernel pada *SVM*. Parameter ini merupakan parameter yang umum digunakan dalam penelitian. Nilai *Complexity C* merupakan parameter yang

terdapat dalam semua kernel *SVM*. Parameter *gamma* digunakan untuk menentukan tingkat kedekatan antara dua titik sehingga memudahkan dalam menemukan pemisah *hyperplane* yang sesuai dengan data. Parameter *degree* berfungsi untuk memudahkan memetakan input data ke ruang dimensi yang lebih tinggi supaya dapat menemukan *hyperlane* yang lebih konsisten, sedangkan parameter *coef* merupakan nilai independen [16].

3.3.8 Evaluasi dan Kesimpulan

Pada proses evaluasi ini, dilakukan analisis hasil klasifikasi dari model dengan data hasil labeling menggunakan *confusion matrix* untuk melihat apakah data hasil klasifikasi dengan data labeling sudah mendapat akurasi, presisi, *recall*, *f1-score* yang baik atau belum. Setelah itu maka dapat ditarik kesimpulan apakah model yang dilatih dapat dengan baik dalam melakukan klasifikasi sentimen terhadap tragedi kanjuruhan atau tidak.