

BAB III METODOLOGI PENELITIAN

3.1. Objek dan Subjek Penelitian



Gambar 3. 1 Video Youtube RANS Entertainment yang dijadikan objek penelitian

Subjek pada studi ini ialah komentar-komentar yang ditinggalkan pada kolom komentar video artis Korea di kanal Youtube Rans Entertainment. Sedangkan objek studi ini ialah video di kanal Youtube RANS Entertainment yang menampilkan artis Korea yang diunggah pada 1 Juni 2023 dan *Support Vector Machine* sebagai algoritma dalam analisis sentimen.

3.2. Alat dan Bahan Penelitian

3.2.1. Alat

Alat yang dipergunakan dalam studi ini dibagi menjadi 2, yakni perangkat keras dan perangkat lunak. Perangkat keras yakni Laptop Dell 64-bit dengan spesifikasi RAM 4 GB. Perangkat lunak yang dipakai terdiri dari Windows 10 dan Google Colab.

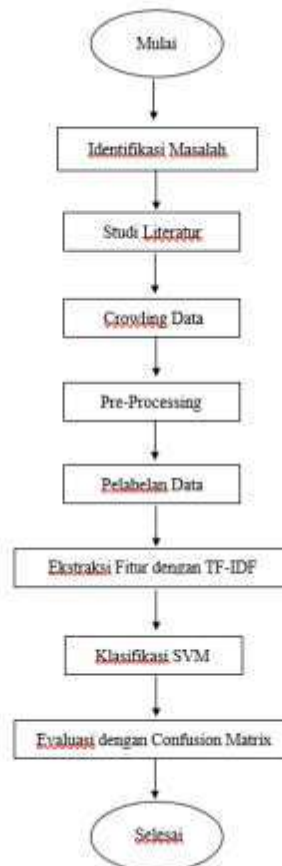
3.2.2. Bahan

Bahan studi ini berupa data komentar yang diambil dari video berjudul “NCT DOJAEJUNG DIAJAK TUMPENGAN DI RUMAH BARU RAFFI NAGITA!!! RAFATHAR & JAEHYUN AKHIRNYA

DIPERTEMUKAN” yang diunggah di kanal Youtube RANS Entertainment pada 1 Juni 2023. Data ini berjumlah komentar 21.757 komentar.

3.3. Diagram Alir Penelitian

Adapun proses studi ini bisa dilihat pada gambar berikut ini.



Gambar 3. 2 *Flowchart* alur penelitian

3.3.1. Identifikasi Masalah

Identifikasi masalah menjadi sesuatu yang sangat mendasar dalam tahapan studi ini. Hal ini dilaksanakan sebagai langkah atau tahapan peneliti menentukan permasalahan atau topik pembahasan, yakni sentiment komentar pada video artis Korea di kanal Youtube RANS Entertainment.

3.3.2. Studi Literatur

Tahapan ini dilaksanakan dengan mencari penelitian dengan topik yang serupa dengan penelitian, contohnya jurnal-jurnal yang membahas topik dan metodologi yang serupa. Studi literatur juga sangat membantu dalam pengolahan data, tahapan pengolahan data dan penggunaan algoritma yang dipergunakan dalam penelitian.

3.3.3. *Crawling Data*

Pengambilan data komentar yang akan dipakai untuk penelitian, diperoleh dari Youtube dengan mempergunakan Youtube API. Data berupa komentar yang ditulis oleh *user* Youtube di video dengan *link* <https://www.youtube.com/watch?v=AOY3F6UzBcY&t=21s>. Proses *crawling* menghasilkan komentar yang kemudian disimpan dalam format .csv. Contoh komentar yang didapatkan dari hasil *crawling* bisa dilihat pada Tabel 3.1

Tabel 3. 1 Contoh Komentar di video artis Korea

No	Contoh komentar
1.	keren juga soalnya memperkenalkan makanan Indonesia ke orang luar 🙌
2.	Aku ialah gambaran raffi ahmad kalau ketemu turis 🤔🤔🤔🤔🤔
3.	Jaehyun terlihat lokal di rumah ini
4.	Bukannya sirik yaa, tapi kok kelakuan mereka norak....
5.	OMG bujang bujang ku akhirnya tumpengan 🤔🤔

3.3.4. *Pre-Processing*

Tahap *pre-processing* dilaksanakan untuk menyeleksi data teks agar menjadi lebih terstruktur lagi dengan melalui serangkaian tahapan yang meliputi *remove duplicated*, *case folding*, *cleansing*, *tokenizing*, *stopword*, dan *stemming*. Contoh dari tahapan *pre-processing* bisa dilihat pada Tabel 3.2, Tabel 3.3, Tabel 3.4, Tabel 3.5, dan Tabel 3.6.

Tabel 3. 2 Contoh hasil *case folding* pada komentar

Teks	Hasil <i>case folding</i>
Luar biasa mbak gigi... budaya Indonesia di perkenalkan	luar biasa mbak gigi... budaya indonesia di perkenalkan

Tabel 3. 3 Contoh hasil *cleansing* pada komentar

Teks	Hasil <i>cleansing</i>
luar biasa mbak gigi... budaya Indonesia di perkenalkan	luar biasa mbak gigi budaya indonesia di perkenalkan

Tabel 3. 4 Contoh hasil *tokenizing* pada komentar

Teks	Hasil <i>tokenizing</i>
luar biasa mbak gigi budaya Indonesia di perkenalkan	luar, biasa, mbak, gigi, budaya, indonesia, perkenalkan

Tabel 3. 5 Contoh hasil *stopword* pada komentar

Teks	Hasil <i>stopword</i>
luar, biasa, mbak, gigi, budaya, Indonesia, perkenalkan	mbak, gigi, budaya, indonesia, perkenalkan

Tabel 3. 6 Contoh hasil *stemming* pada komentar

Teks	Hasil <i>stemming</i>
mbak, gigi, budaya, indonesia, perkenalkan	mbak gigi budaya indonesia kenal

3.3.5. Labeling

Tahap *labeling* bertujuan untuk memberi nilai/sentimen pada komentar. Proses *labeling* pada studi ini dilaksanakan dengan mempergunakan kamus *lexicon*. Kamus *lexicon* mengevaluasi teks dalam bentuk daftar kata untuk menentukan sentimennya. Kamus *lexicon* berisi daftar kata positif dan negatif dengan nilai polaritasnya.

Metode ini melaksanakan kategori data dengan memperbandingkan jumlah kata positif dan jumlah kata negatif pada suatu dokumen.

Penelitian ini menggunakan kamus Inset Lexicon yang dikemukakan oleh Fajri (2017) dalam penelitian yang berjudul “Evaluasi daftar kata untuk sentimen analisis berbahasa Indonesia”, terdiri dari 3.609 kata pada kamus *lexicon* positif dan 6.609 kata pada kamus *lexicon* negatif. Studi ini membagi data menjadi tiga sentimen yakni positif, negatif, atau netral. Jika skor polariti lebih dari 0 maka komentar dilabeli positif, jika 0 maka netral dan jika kurang dari 0 merupakan negatif. Contoh hasil *labeling* bisa dilihat dalam tabel berikut ini.

Tabel 3. 7 Contoh *labeling* pada komentar dengan kamus *lexicon*

No	Teks komentar	Teks <i>pre-processing</i>	Skor polariti	Label Sentimen
1.	Doyoung antusias banget liat tumpeng	doyoung, antusias, banget, lihat, tumpeng	3	positif
2.	Siapa yang tiap hari nonton ini?? Saya saya	nonton	0	netral
3.	Itu gimana sih penerjemahnya diem aja gak jelasin ini itu, ngejelasinnya pas disuruh doang	terjemah, diam, jelasin, ngejelasinnya, suruh	-4	negatif

Untuk perhitungan hasil skor polariti dari teks komentar di atas dapat dilihat pada tabel di bawah ini.

Tabel 3. 8 Contoh perhitungan hasil skor polariti pada komentar

Ko men tar	Kata pada teks <i>pre- processing</i>	Lexicon Positif	Lexicon Negatif	Skor
1	Doyoung	-	-	0
	Antusias	Tercantum (2)	Tercantum (-3)	-1
	Banget	Tercantum (1)	-	1
	Lihat	Tercantum (3)	-	3
	Tumpeng	-	-	0
	Total Skor			3 (Sentimen Positif)
2	Nonton	-	-	0
	Total Skor			0 (Sentimen Netral)
3	Terjemah	-	-	0
	Diam	Tercantum (1)	Tercantum (- 3)	-2
	Ngejelasinnya	-	-	0
	Suruh	-	Tercantum (- 2)	-2
	Total Skor			-4 (Sentimen Negatif)

3.3.6. TF-IDF

Tahapan studi ini melaksanakan pembobotan mempergunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Pembobotan atau penilaian dilaksanakan berdasarkan frekuensi kemunculan kata pada

dokumen. Untuk contoh perhitungan pembobotan TF-IDF bisa dilihat pada Tabel 3.9, Tabel 3.10, dan Tabel 3.11 berikut ini.

Tabel 3. 9 Contoh pembobotan TF-IDF pada komentar Youtube

Dok	Komentar
1	terima kasih papa rafi mamah gigi undang dojaejung
2	lihat dojaejung main perosot andara
3	lihat jaehyun salim rafathar

Proses perhitungan TF-IDF dimulai dengan menghitung nilai TF (Term Frequency). Adapun proses perhitungannya mengacu pada persamaan 2.1.

Tabel 3. 10 Contoh perhitungan TF untuk dokumen pada Tabel 3.9

TF	TF			
Term	D1	D2	D3	DF
terima	1	0	0	1
kasih	1	0	0	1
papa	1	0	0	1
raffi	1	0	0	1
mamah	1	0	0	1
gigi	1	0	0	1
undang	1	0	0	1
dojaejung	1	1	0	2
lihat	0	1	1	2
main	0	1	0	1
perosotan	0	1	0	1
andara	0	1	0	1
jaehyun	0	0	1	1
salim	0	0	1	1
rafathar	0	0	1	1

Proses selanjutnya adalah menghitung nilai IDF dan TF-IDF. Adapun proses perhitungannya mengacu pada persamaan 2.2.

Tabel 3. 11 Contoh perhitungan TF-IDF untuk dokumen pada Tabel 3.9

Term	IDF	TF IDF		
		D1	D2	D3
terima	$\text{Log}(3/1) = 0,477$	0,477	0	0
kasih	$\text{Log}(3/1) = 0,477$	0,477	0	0
papa	$\text{Log}(3/1) = 0,477$	0,477	0	0
raffi	$\text{Log}(3/1) = 0,477$	0,477	0	0
mamah	$\text{Log}(3/1) = 0,477$	0,477	0	0
gigi	$\text{Log}(3/1) = 0,477$	0,477	0	0
undang	$\text{Log}(3/1) = 0,477$	0,477	0	0
dojaejung	$\text{Log}(3/2) = 0,176$	0,176	0,176	0
lihat	$\text{Log}(3/2) = 0,176$	0	0,176	0,176
main	$\text{Log}(3/1) = 0,477$	0	0,477	0
perosotan	$\text{Log}(3/1) = 0,477$	0	0,477	0
andara	$\text{Log}(3/1) = 0,477$	0	0,477	0
jaehyun	$\text{Log}(3/1) = 0,477$	0	0	0,477
salim	$\text{Log}(3/1) = 0,477$	0	0	0,477
rafathar	$\text{Log}(3/1) = 0,477$	0	0	0,477

3.3.7. Klasifikasi *Support Vector Machine*

Proses klasifikasi pada analisis ini dilaksanakan dengan mempergunakan *Support Vector Machine*. *Support vector machine* ialah salah satu algoritma klasifikasi yang sering dipergunakan untuk analisis sentimen. Tujuan dari metode *Support vector machine* ialah menemukan *hyperplane* yang paling optimum. *Hyperplane* yang optimum ialah sebuah *hyperplane* yang jika tepat berada di tengah-tengah kelas sehingga mempunyai jarak paling jauh ke data-data terluar di masing-masing kelas.

3.3.8. *Confusion Matrix*

Setelah melaksanakan proses pembobotan kata dan klasifikasi mempergunakan *Support Vector Machine*, maka dilaksanakan tahap pengujian untuk mencari nilai akurasi, presisi, *recall*, dan *f1-score*. Pengujian ini akan memperbandingkan data *testing* yang sudah berlabel kelas positif, kelas negatif, dan kelas netral mempergunakan *confusion matrix*.