

## BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

### 2.1. Tinjauan Pustaka

Analisis sentimen ialah suatu metode guna melihat pendapat warga pada subjek tersendiri dari suatu perkumpulan informasi. Seiring dengan bertumbuhnya informasi yang makin canggih berdampak pada berubahnya manusia terkait komunikasi dengan sesama. Penggunaan media sosial banyak dipergunakan oleh warga umum teruntuk komunikasi ataupun berekspresi mengutarakan pendapatnya. Studi terkait analisa sentimen hingga saat ini sudah banyak dilaksanakan [8].

Penelitian pertama berjudul “Analisis Sentimen Komentar *Youtube TvOne* Terkait Ustadz Abdul Somad Dideportasi Dari Singapura Mempergunakan *Algoritma SVM*” ditulis oleh Desi Mualfah, Ramadhoni, Rahmad Gunawan, dan Danang Mulyadipa Suratno pada 2023. Masalah studi ini ialah video terkait Ustadz Abdul Somad di deportasi dari Singapura mendapat banyak *feedback* yakni komentar positif ataupun negatif dari warga dan membutuhkan waktu lama jika diklasifikasi secara manual, oleh karena itu perlu dilaksanakan klasifikasi mempergunakan *machine learning*. Studi ini mempergunakan algoritma *Support Vector Machine* dan menghasilkan hasil akurasi yang tinggi yakni 95,02% [5].

Penelitian kedua berjudul Analisis Sentimen Pada Komentar Video Youtube Mempergunakan *Support Vector Machine* oleh Toif Muhayat, Ahmad Fauzi, dan Jamaludin Indra. Studi ini bertujuan agar pembuat konten Youtube bisa mengetahui jenis video yang diminati oleh penonton. Metode yang dipergunakan ialah algoritma *Support Vector Machine* dengan ekstraksi fitur TF-IDF. Keluaran hasil studi ini berupa jenis konten *daily vlog* mempunyai sentimen dominan positif dengan persentase 84% sentimen positif dan 16% sentimen negatif. Sementara akurasi yang dihasilkan 86%, presisi 87%, *recall* 99%, dan *f1-score* 100% [9].

Penelitian ketiga dilaksanakan oleh Ardhi Wijayanto dan Afrima Dhia Defara (2022) dengan judul “ Analisis Sentimen Komentar Youtube Terkait Vaksin Covid-19 Mempergunakan *Support Vector Machine*”. Masalah dalam studi ini ialah kebijakan pemerintah terutama terkait kebijakan vaksinasi tidak selalu bisa diterima oleh warga. Untuk lebih mengerti sentimen warga terkait aturan vaksin Covid-19, maka dilaksanakan analisa pada komentar Youtube mempergunakan algoritma *Support Vector Machine*. Studi ini mempergunakan 1.007 komentar dan 3 skenario rasio data *training* dan data *testing* yang berbeda. Pengujian algoritma *Support Vector Machine* pada studi ini menghasilkan akurasi rata-rata 85,6%, *recall* 90,1%, *f1-score* 84,7%, dan presisi 80,1% dengan rasio terbaik yakni 90:10 [10].

Penelitian keempat “Analisis Sentimen Twitter Pada Penggunaan Artis Korea Selatan sebagai Brand Ambassador Produk Kecantikan Lokal” pada 2022 dilaksanakan oleh Ristyanti Slamet, Windu Gata, Annisa Novtariany, Khairunnisa Hilyati, Febri Ainun Jariyah. Masalah studi ini mengacu pada pro dan kontra terkait penggunaan artis asal Korea Selatan sebagai *brand ambassador* produk kecantikan Indonesia dari sudut pandang *user brand* kecantikan maupun penggemar artis Korea. Studi ini memperbandingkan algoritma *Naïve Bayes* dan *Support Vector Machine*. Data penelitian dari Twitter dengan mempergunakan kata kunci “skincare” dan “BA”. Hasil penelitian menunjukkan akurasi *Naïve Bayes* sejumlah 82,65% dan SVM sejumlah 83,60% [11].

Penelitian kelima “Perbandingan *Naïve Bayes*, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek” ditulis oleh Jessica Widyadhana Iskandar dan Yessica Nataliani. Tujuan studi ini untuk mengetahui sentimen dari komentar Youtube terkait produk *gadget* Samsung Galaxy Z Flip 3. Studi ini mempergunakan model CRISP-DM dan memperbandingkan 3 metode kategori yakni *Naïve Bayes*, *Support Vector Machine* dan *K-Nearest Neighbor*. Hasil penelitian menunjukkan, *Support Vector Machine* mempunyai rata-rata akurasi terbaik sejumlah 96,43% [12]

Tabel 2. 1 Penelitian Terkait

No.	Penulis	Judul Penelitian	Tahun	Metode	Masalah	Hasil
1.	Desi Mualfah, Ramadhoni, Rahmadi, Gunawan, Danang Mulyadipa Suratno	Analisis Sentimen Komentar <i>Youtube TvOne</i> Terkait Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma <i>SVM</i>	2023	Algoritma <i>Support Vector Machine</i>	<i>Feedback</i> komentar maupun negatif dari warga terkait video Youtube Ustadz Abdul Somad dideportasi dari Singapura membutuhkan waktu lama jika diklasifikasi secara manual, oleh karena itu perlu dilaksanakan klasifikasi menggunakan <i>machine learning</i>	Akurasi yang dihasilkan <i>Support Vector Machine</i> 95,02%

No.	Penulis	Judul Penelitian	Tahun	Metode	Masalah	Hasil
2.	Toif Muhayat, Ahmad Fauzi, Jamaludin Indra	Analisis Sentimen Pada Komentar Video Youtube Menggunakan Support Vector Machine	2023	algoritma <i>Support Vector Machine</i> dengan ekstraksi fitur TF-IDF	Mengetahui jenis video yang diminati oleh penonton dari analisis sentimen komentar Youtube	Jenis konten <i>daily vlog</i> mempunyai sentimen dominan positif dengan persentase 84% sentimen positif dan 16% sentimen negatif. serta akurasi yang dihasilkan 86%, presisi 87%, <i>recall</i> 99%, dan <i>f1-score</i> 100%
3.	Ardhi Wijayanto. Afrima	Analisis Sentimen Komentar Youtube	2022	Algoritma <i>Support Vector Machine</i> dengan 3 skenario rasio data training	Memahami sentimen warga terkait kebijakan vaksinasi Covid-19	<i>Support Machine</i> menghasilkan akurasi rata-rata

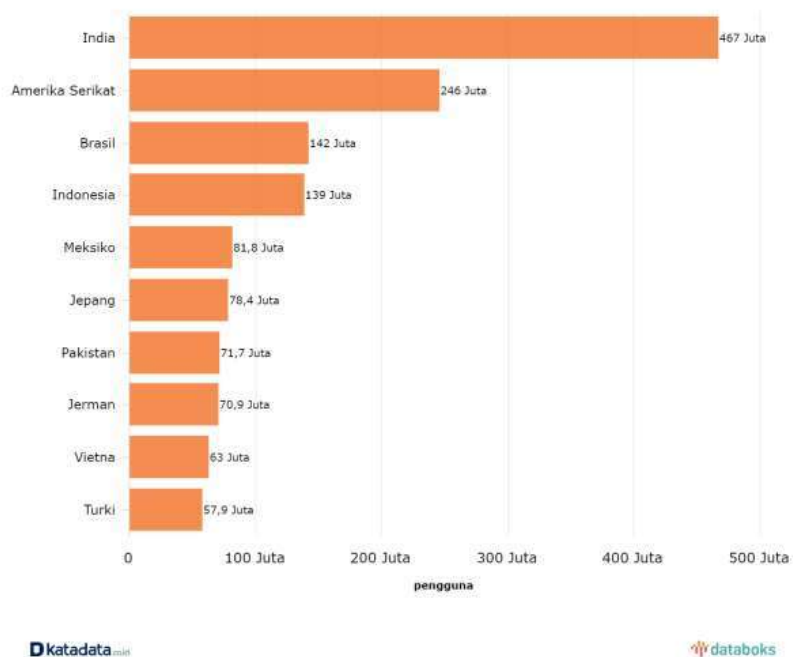
No.	Penulis	Judul Penelitian	Tahun	Metode	Masalah	Hasil
	Dhia Defara	Terkait Vaksin Covid-19 Menggunakan <i>Support Vector Machine</i>		dan data testing yang berbeda		85,6%, <i>recall</i> 90,1%, <i>f1-score</i> 84,7%, dan presisi 80,1% dengan rasio terbaik yakni 90:10
4.	Ristyanti Slamet, Windu Gata, Annisa Novtariany, Khairunni sa Hilyati, Febri	Analisis Sentimen Twitter Pada Penggunaan Artis Korea Selatan sebagai <i>Brand Ambassador</i> Produk Kecantikan Lokal	2022	Algoritma Naive Bayes dan <i>Support Vector Machine</i>	Penggunaan artis asal Korea Selatan sebagai <i>brand ambassador</i> produk kecantikan Indonesia menuai pro dan kontra. Hal itu menyebabkan penelitian dilaksanakan dengan menganalisis sentimen dari komentar <i>user</i>	Hasil penelitian menunjukkan akurasi Naive Bayes sejumlah 82,65% dan SVM sejumlah 83,60%

No.	Penulis	Judul Penelitian	Tahun	Metode	Masalah	Hasil
	Ainun Jariyah				<i>brand</i> kecantikan maupun penggemar artis Korea di Twitter	
5.	Jessica Widyadhana Iskandar, Yessica Nataliani	Perbandingan Naive Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek	2021	Mempergunakan model CRISP-DM dan 3 algoritma yakni Naive Bayes, <i>Support Vector Machine</i> dan K-Nearest Neighbor	Mengetahui sentimen dari komentar Youtube terkait produk gadget Samsung Galaxy Z Flip 3	<i>Support Vector Machine</i> mempunyai rata-rata akurasi terbaik sejumlah 96,43%

Berdasarkan Tabel 2.1 yang memuat penelitian terdahulu diketahui, analisis sentimen yang dilaksanakan dengan algoritma *Support Vector Machine* menghasilkan skor akurasi tinggi, sehingga algoritma *Support Vector Machine* akan diterapkan dalam studi ini. Studi kasus yang dipergunakan pada penelitian tugas akhir ini ialah yang membedakannya dengan riset-riset terdahulu.

## 2.2. Landasan Teori

### 2.2.1. Youtube



Gambar 2.1 Negara dengan *user* Youtube terbanyak di dunia  
Sumber: We Are Social, 2023

Youtube saat ini sudah menjadi situs media terpopuler dan menempati ranking kedua di dunia sebagai media sosial dengan *user* terbanyak. Menurut survei yang dilaksanakan oleh We Are Social dan Hootsuite, per Januari 2023 *user* Youtube mencapai 2,51 miliar. Indonesia sendiri menyumbang 139 juta *user* dan menjadikannya berada di peringkat keempat dari seluruh dunia [13].

Youtube ialah suatu situs web serta aplikasi yang memudahkan *user* untuk mengirimkan video atau menyiarkan video langsung yang bisa dilihat oleh *user* lain tanpa bayar dimanapun [14]. Sederhananya,

Youtube menyediakan *platform* yang tampaknya gratis kepada para pembuat konten, yang kemudian saling berlomba-lomba menarik perhatian penonton yang cukup besar lewat konten masing-masing. Youtube lalu menjual konten tersebut kepada para pengiklan yang terus naik dari tahun-tahun sebelumnya [1].

### 2.2.2. RANS Entertainment

Raffi Ahmad dan Nagita Slavina, pasangan selebriti Indonesia, mendirikan RANS Entertainment pada 27 Desember 2015. Nama Raffi Ahmad (RA) dan Nagita Slavina (NS) digabungkan menjadi nama RANS. RANS Entertainment awalnya hanya sebuah saluran Youtube, namun kini berkembang menjadi sebuah badan usaha hiburan dan rumah produksi.



Gambar 2.2 Data statistik Youtube RANS Entertainment  
Sumber: Socialblade, 2023

RANS Entertainment sendiri menempati urutan kelima sebagai akun Youtube dengan jumlah *subscriber* terbanyak di Indonesia [4]. Terhitung pada 25 Juli 2023, kanal Youtube RANS Entertainment berhasil mempunyai 25,4 juta *subscriber*. Saluran ini sudah mengunggah 3.719 video dengan total penonton sejumlah 6,4 miliar.

Tidak hanya berfokus pada dunia hiburan, RANS Entertainment kini melebarkan sayap ke beragam industri contohnya olahraga,



kecantikan hingga properti. Dilansir dari situs katadata.co.id, RANS Entertainment mempunyai 10 lini bisnis dimana 8 sudah beroperasi dan 2 belum beroperasi, yakni RANS Music, PowerRANgers, RA Picture, RANS Animation Studio, RANS E-Sports, RANS Nusantara FC, RANS PIK Basketball Club, RANS Beauty, RANS Karnaval International, dan Apatemen Andara Living [15].

### 2.2.3. NCT Dojaejung

NCT Dojaejung ialah salah satu sub unit grup NCT yang bernaung di SM Entertainment [16]. Nama Dojaejung sendiri ialah gabungan dari nama ketiga anggotanya yakni Doyoung, Jaehyun dan Jungwoo. Sebelum bergabung dengan NCT Dojaejung, Doyoung, Jaehyun dan Jungwoo sudah aktif lebih dulu di NCT 127. Sub unit ini debut tanggal 17 April 2023 merilis mini album *Perfume*. NCT Dojaejung berhasil menjual album debut mereka dengan angka yang terbilang besar yakni 845.978 eksemplar.

### 2.2.4. Analisis Sentimen

Penggunaan media sosial yang berkembang pesat, saat ini menawarkan kumpulan data yang sangat berharga. Akan tetapi, data-data yang tersedia di media sosial ini bersifat tidak terstruktur, multimodal, informal, berdimensi tinggi, dan beragam. Analisis sentimen dari pendapat *user* ialah pendekatan yang baik yang dipergunakan untuk menganalisis data dan mengklasifikasinya menjadi positif, negatif atau netral [17].

Analisa sentimen ialah tahapan mencari arti suatu sudut pandang, opini ataupun ekspresi yang ada dalam suatu teks, pernyataan, *post* (kegiatan di internet), ataupun basis informasi dengan pendekatan *Natural Language Processing* [18].

### 2.2.5. Pre-processing

*Pre-processing* ialah tahapan sebelum data dilaksanakan lebih lanjut. Tahap *pre-processing* ini bertujuan untuk mempersiapkan data

teks sehingga menghasilkan data teks dengan kualitas baik. Tahapan *pre-processing* data dilaksanakan dengan beragam langkah berikut [19]:

1. *Remove Duplicated*

Tahapan ini dilaksanakan bertujuan mengurangi kesalahan dalam teks dokumen dengan melaksanakan penghapusan data yang dinilai sama.

2. *Cleansing*

Tahapan ini ialah tahapan pembersihan teks dari kata ataupun karakter-karakter yang tidak diperlukan sehingga topik penelitian lebih mudah difokuskan.

3. *Case Folding*

Tahapan ini dilaksanakan dengan mengubah bentuk huruf kapital pada data teks menjadi huruf kecil.

4. *Tokenizing*

Tahapan ini ialah proses pemisahan kata demi kata menjadi potongan kata yang nantinya disebut sebagai atribut.

5. *Stemming*

Tahapan ini mengubah atribut menjadi *root words* atau kata-kata dasar dengan menghilangkan kata imbuhan.

6. *Stopword*

Tahapan ini ialah tahapan penyaringan kata dengan melaksanakan penghapusan kata yang tidak diperlukan sehingga menghasilkan kata-kata yang dianggap bisa mewakili data teks.

#### 2.2.6. TF-IDF

TF-IDF ialah kependekan dari *Term Frequency-Inverse Document Frequency*, sebuah teknik mengkalkulasi bobot tiap kata pada komentar, baik data *training* maupun data *testing*. Teknik ini mengkalkulasi skor TF serta IDF di tiap token di tiap dokumen. TF ialah kuantitas munculnya suatu *term* dalam suatu dokumen [20]. Sedangkan IDF ialah pemberian skor kata berdasarkan jumlah dokumen yang mencakup suatu kata (DF). Makin banyak dokumen dengan kata spesifik,

makin kecil pengaruh kata tersebut terhadap dokumen. Tapi, makin sedikit dokumen dengan suatu kata spesifik, makin besar dampaknya terhadap dokumen [21]. *Skoring* mempergunakan TF-IDF dihitung dengan mempergunakan persamaan berikut [22]:

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (2.1)$$

$$TF - IDF(w, d) = TF(w, d) \times IDF(w) \quad (2.2)$$

Dimana:

- $TF - IDF(w, d)$  : bobot suatu kata didalam seluruh dokumen
- $w$  : suatu kata (*word*)
- $d$  : suatu dokumen (*document*)
- $TF(w, d)$  : jumlah munculnya suatu kata  $w$  didalam dokumen
- $IDF(w)$  : *inverse* DF dari kata  $w$
- $N$  : kuantitas seluruh dokumen

#### 2.2.7. Support Vector Machine

*Support Vector Machine* ialah sebagian dari *supervised learning* yang dikemukakan oleh Vapnik guna mendukung pengkategorian pada beragam objek dalam beragam domain[23]. Misalnya melaksanakan kategori gambar, dan teknik identifikasi objek lainnya contohnya tanda tangan[24].

Algoritma *Support Vector Machine* bisa melaksanakan kategori untuk beragam kelas, yang sering dikenal sebagai *multiclass*, termasuk ketika ingin mengkategorikan *text mining* dengan mempergunakan tiga sentimen, yakni ada negatif positif, serta netral. Ada beragam pendekatan yang bisa dipergunakan dalam *Support Vector Machine* untuk membantu mengklasifikasikan tiga kelas atau lebih, yakni [23]:

### 1. *One Againsts One*

Pendekatan ini membuat satu pengklasifikasi untuk tiap pasangan kelas. Jadi jumlah pengklasifikasi bisa diketahui dari jumlah kelas di mana  $N$  ialah jumlah kelas. Kemudian, pengklasifikasi dilatih untuk membedakan sampel dari satu kelas dengan sampel dari kelas lain. *Maximum vote* akan dipergunakan untuk memperkirakan kategorisasi pola yang tidak teridentifikasi. Tiap *classifier* akan menghasilkan satu suara dan suara tersebut kemudian akan dihitung untuk menentukan kelasnya .

### 2. *One Againsts All*

Dalam pendekatan *One Againsts All* mempunyai satu *classifier* per kelas yang kemudian dilatih untuk membedakan sampel dari satu kelas dengan sampel dari semua kelas lainnya. Pola yang tidak diketahui biasanya diklasifikasikan berdasarkan *output* maksimum gabungan *Support Vector Machine*.

#### 2.2.8. *Confusion Matrix*

Untuk mengidentifikasi nilai dengan tingkat akurasi tertinggi, kategori analisis sentiment akan mengkalkulasi skor *recall*, presisi, akurasi, serta *f1-score* [25]. Pada Tabel 2.2 merupakan tabel *confusion matrix* dengan tiga kelas, yaitu negatif, netral dan positif.

Tabel 2. 2 *Confusion matrix* tiga kelas

Aktual	Kelas Negatif (Prediksi)	Kelas Netral (Prediksi)	Kelas Positif (Prediksi)
Negatif	<i>True Negative</i> (TNg)	<i>False Neutral1</i> (FNt1)	<i>False Positive1</i> (FP1)
Netral	<i>False Negative1</i> (FNg1)	<i>True Neutral</i> (TNt)	<i>False Positive2</i> (FP2)
Positif	<i>False Negative2</i> (FNg2)	<i>False Neutral2</i> (FNt2)	<i>True Positive</i> (TP)

Untuk persamaan matriks tiga kelas bisa dilihat dalam keterangan berikut ini.

1. Akurasi, dengan tujuan untuk mengidentifikasi kategori yang akurat.

Rumus akurasinya adalah:

$$Akurasi = \frac{(TNg + TNt + TP)}{(TNg + FNt1 + \dots + FNt2 + TP)} \times 100\% \quad (2.3)$$

2. Presisi, dipergunakan untuk mengidentifikasi skor dari kelas yang diberikan, contohnya mengidentifikasi skor positif didalam kelas positif yang sudah dikategorikan. Persamaan presisi sebagai berikut:

$$Presisi\ Positif = \frac{TP}{(TP + FP1 + FP2)} \times 100\% \quad (2.4)$$

$$Presisi\ Negatif = \frac{TNg}{(TNg + FNg1 + FNg2)} \times 100\% \quad (2.5)$$

$$Presisi\ Netral = \frac{TNt}{(TNt + FNt1 + FNt2)} \times 100\% \quad (2.6)$$

3. *Recall*, memperkirakan skor seluruh dokumen dari teks. Misalnya, mempergunakan semua data positif untuk memperbandingkan temuan sentimen positif. Persamaan *recall* yakni:

$$Recall\ Positif = \frac{TP}{(TP + FNg2 + FNt2)} \times 100\% \quad (2.7)$$

$$Recall\ Negatif = \frac{TNg}{(TNg + FNt1 + FP1)} \times 100\% \quad (2.8)$$

$$Recall\ Netral = \frac{TNt}{(TNt + FNg1 + FP2)} \times 100\% \quad (2.9)$$

4. *F1-score*, bisa ditentukan berdasarkan bobot perbandingan, ialah rata-rata dari presisi dan *recall*.

$$f1 - score = \frac{2(Presisi \times Recall)}{(Recall + Presisi)}$$

(2. 10)

Dimana:

TP	: <i>True Positive</i>
TNg	: <i>True Negative</i>
TNt	: <i>True Neutral</i>
FP1	: <i>False Positive1</i>
FP2	: <i>False Positive2</i>
FNg1	: <i>False Negative1</i>
FNg2	: <i>False Negative2</i>
FNt1	: <i>False Neutral1</i>
FNt2	: <i>False Neutral2</i>

Hasilnya akan diklasifikasikan berdasarkan tingkat akurasi yang dicapai. Berikut ini adalah rentang akurasi yang digunakan [26]:

1. 0.90-1.00 = *Excellent classification*
2. 0.80-0.90 = *Good classification*
3. 0.70-0.80 = *Fair classification*
4. 0.60-0.70 = *Poor classification*
5. 0.50-0.60 = *Failure*

## 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