

Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine

Okta Ihza Gifari¹, Muh. Adha², Ivan Rifky Hendrawan³, Fernandito Freddy Setlight Durrand⁴

^{1,2,3,4}Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta Jalan Ring Road Utara Condong Catur Depok-Sleman

¹oktagifari@students.amikom.ac.id, ²muh.adha@students.amikom.ac.id, ³ivanrifky@students.amikom.ac.id,

⁴nandodurand@students.amikom.ac.id

Abstract— With today's technological advancements, all information about all movies is readily available on the Internet. If information is managed properly, it can provide benefits in the form of useful information to help individuals or organizations to make decisions. This study aims to explain sentiment analysis on film documents. The methods used in this research are TF-IDF (Term Frequency-Inverse Document Frequency) and SVM (Support Vector Machine). This method was chosen because it is capable of weighting words and classifying high-dimensional data. From the scenario tests conducted, it is known that the TF-IDF and SVM algorithms can be used for film review cases with an Accuracy value of 85%, a Precision value of 100%, a Recall value of 70%, and an F1-Score value of 82%.

Intisari— Dengan kemajuan teknologi saat ini seluruh informasi tentang semua film sudah tersedia di Internet. Jika informasi dikelola dengan baik maka dapat memberikan manfaat berupa informasi yang berguna untuk membantu individu atau organisasi untuk mengambil keputusan. Penelitian ini bertujuan untuk menjelaskan analisis sentimen pada dokumen film. Metode yang digunakan pada penelitian ini adalah TF-IDF (Term Frequency-Inverse Document Frequency) dan SVM (Support Vector Machine). Metode ini dipilih karena mampu melakukan pembobotan terhadap kata dan mengklasifikasikan data berdimensi tinggi. Dari uji skenario yang dilakukan, diketahui bahwa algoritma TF-IDF dan SVM dapat digunakan untuk kasus review film dengan nilai Accuracy 85%, nilai Precision 100%, nilai Recall 70%, dan nilai F1-Score sebesar 82%.

Kata Kunci— analisis sentimen, Term Frequency-Inverse Document Frequency, Support Vector Machine, review film, klasifikasi

I. PENDAHULUAN

Kebutuhan akan informasi didorong oleh sejumlah besar penelitian dan teknologi. Informasi yang kita miliki saat ini begitu luas dan beragam, sehingga dibutuhkan berbagai cara penyajian informasi untuk mencakup keseluruhannya [1]. Setiap pengguna internet aktif memiliki kebebasan untuk mengungkapkan pendapat mereka tentang topik tertentu. Salah satu contoh jenis resensi adalah resensi film.

Banyak dokumen teks berasal dari berbagai sumber, seperti review, opini, informasi, artikel, buku, perpustakaan digital, pesan email, dan halaman web. Kemudahan penggunaan website mengakibatkan bertambahnya dokumen teks yang berupa pendapat dan opini. Teknik yang berkembang saat penggalan dokumen teks waktu ini adalah text mining.

Penambangan dokumen teks dari website yang berisi komentar, pendapat, feedback, kritik dan review merupakan hal penting, karena apabila dikelola dengan baik maka dapat memberikan manfaat berupa informasi yang berguna untuk membantu individu atau organisasi untuk mengambil keputusan. Text Mining merupakan suatu proses pengambilan intisari berasal dokumen teks sehingga didapatkan hasil yang bermanfaat untuk tujuan tertentu [2].

Kemudian sesuai menggunakan buku The Text Mining Handbook [3], text mining dapat didefinisikan menjadi suatu proses menggali informasi dimana seorang user berinteraksi menggunakan sekumpulan dokumen memakai tools analisis yg merupakan komponen-komponen dalam data mining yang salah satunya merupakan dengan klasifikasi. Teknologi text mining mencakup beberapa kategori. Salah satunya adalah analisis sentimen, proses otomatis memahami, mengekstraksi, dan memanipulasi data tekstual, atau penelitian komputerisasi opini, emosi, dan emosi yang diungkapkan dalam format teks. Ada beberapa algoritma atau metode untuk analisis sentimen, antara lain Support Vector Machine (SVM) dan term frequency-inverse document frequency.

Di dalam penelitian ini algoritma yang digunakan adalah TF-IDF sebagai pembobotan kata dan Support Vector Machine (SVM) untuk pembagian data menjadi dua kelas.

II. TINJAUAN PUSTAKA

A. Studi Literature

Dalam penelitian ini, beberapa literatur pencarian dilakukan untuk menemukan studi relevan yang dilakukan, termasuk studi oleh [4] yang bertujuan untuk menjelaskan klasifikasi sentimen dokumen ulasan film. Metode yang dipergunakan pada penyelesaian kasus peneliti berikut ini menggunakan Support Vector Machine (SVM). Pengujian skenario menunjukkan bahwa algoritma SVM dapat digunakan untuk menganalisis kasus ulasan film dengan Skor F1 sebesar 84,9%.

Tweet tentang vaksin Covid-19 menuai berbagai opini positif dan negatif di masyarakat. Oleh karena itu, penting untuk mengidentifikasi dan mengecualikan informasi yang salah. Sebuah studi oleh [5] tentang analisis sentimen tweet vaksin Covid-19 yang menyelidiki metode Recurrent Neural Network (RNN) dan naive bayes. Data record (dataset) yang dipergunakan digunakan mulai dari 5.000 tweet vaksinasi Covid-19, serta 3800 tweet sentimen positif, 800 tweet sentimen negatif, dan 400 tweet sentimen bebas. Penelitian ini menggunakan metode TF IDF untuk pembobotan dokumen. Hasil pengujian menunjukkan bahwa akurasi RNN (TF IDF) adalah 97,77% lebih tinggi dari 80% untuk Naive Bayes.

Penelitian selanjutnya oleh [6] bertujuan untuk membangun sistem pendeteksi rumor Indonesia di Twitter menggunakan klasifikasi SVM dan pemilihan fitur dengan pembobotan TF IDF. Dataset yang digunakan adalah metode crawl dan proses manual labelling. Subjek data yang digunakan adalah tren sekitar dengan 47.499 dataset. Langkah selanjutnya adalah pemilihan fitur, pemodelan N-Gram, dan evaluasi matriks konfusi. Berdasarkan hasil investigasi, sistem bekerja dengan baik dengan nilai akurasi 78,71%.

B. Analisis Sentiment

Analisis sentimen adalah teknik mengetahui, mengekstrak, dan menjalankan informasi tekstual secara otomatis untuk menemui keterangan sentimental dalam mengungkapkan pemikiran. Sentiment analysis dipergunakan untuk melihat pendapat atau kesamaan opini terhadap sebuah persoalan atau objek oleh seorang menuju ke opini positif atau negatif [7].

C. Text Mining

Text mining merupakan proses ekstraksi berita dari data asal yang belum terstruktur. Data yang belum terstruktur akan diolah memakai teknik serta metode tertentu membuat berita yang bermanfaat untuk pengguna. Text mining adalah teknik yang dipergunakan untuk menangani persoalan classification, clustering, information extraction, serta information retrieval [8].

D. Preprocessing

Preprocessing adalah proses untuk menormalkan istilah – istilah berasal kalimat. Hal ini dilakukan untuk menerima data latih yang baik serta fitur – fitur yang diekstrak nantinya sinkron dengan yang diinginkan. sehingga mampu menyederhanakan pemrosesan data. Mengumpulkan data opini dari media sosial Twitter tidak boleh identik dengan kata baku, kata yang tidak ada dalam kamus, atau bahasa daerah yang digunakan atau dihilangkan. Untuk mengembalikan sejumlah teks ke teks alami dengan mengeliminasi ekspresi atipikal agar dapat meminimalkan noise pada tahap selanjutnya, diperlukan pre-processing atau normalisasi untuk mengatasi hal ini. Menurut contoh pada buku [9] Pra Pengolahan terdapat 6 proses yang terdiri dari :

1) Normalization

Tahap ini dilakukan agar kata-kata yang disingkat atau diperpanjang menjadi kata-kata yang normal sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Konversi slang word adalah proses mengubah kata tidak baku menjadi kata baku.

2) Cleansing

Tahap ini adalah tahap eliminasi aksara non alfabetis untuk menurunkan noise. Aksara yang dihapus adalah tanda baca seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!), serta simbol-simbol seperti tanda '@' untuk username, hashtag (#), emoticon, dan alamat website

3) Case Folding

Case folding adalah tahap untuk mengkonversi karakter alfabet yang telah melalui tahap cleansing ke huruf kecil (lower case).

4) Tokenizing

Tahap ini berfungsi sebagai pemecah kalimat berdasarkan tiap kata yang menyusunnya, yang disebut term atau token. Tokenizing dipecah berdasarkan spasi.

5) Filtering

Tahap ini memproses agar kata-kata yang tidak penting atau tidak bermakna dihapus untuk analisis sentimen. Contoh kata-kata tersebut adalah atau, yang, dengan, di, ke, dan tetapi.

6) Stemming

Tahap ini berfungsi mengubah kata berimbuhan pada tiap kata yang telah terseleksi menjadi kata dasar [9]

E. Term Frequency-Inverse Document Frequency

Term weighting adalah suatu proses pembobotan setiap kata agar bisa mengoptimalkan kemampuan analisis sentimen pada proses text mining. Penelitian ini memanfaatkan Term Frequency-Inverse Document Frequency (TF-IDF).

Term Frequency (tf(w,d)) dianggap memiliki proporsi kepentingan sesuai total kemunculannya dalam teks atau dokumen. Inverse Document Frequency (IDF) merupakan metode pembobotan token yang berfungsi untuk memonitor kemunculan token dalam himpunan teks.

F. Support Vector Machine

SVM adalah teknik untuk memprediksi klasifikasi dan regresi [6]. SVM pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai seperangkat konsep unggul yang di bidang pattern recognition[7]. SVM adalah algoritma pembelajaran mesin yang bekerja berdasarkan prinsip Structural Risk Minimization (SRM) dengan tujuan untuk menemukan hyperplane optimal yang memisahkan dua kelas dalam ruang input.

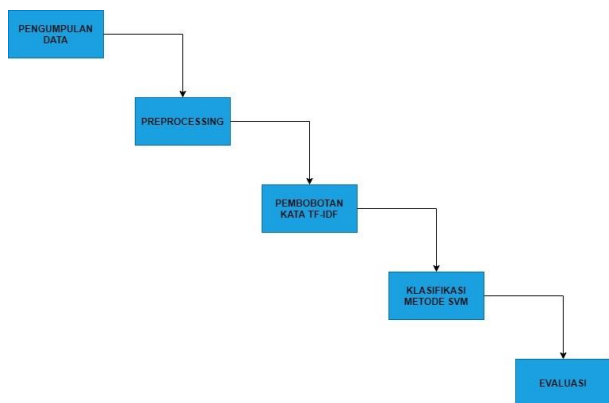
G. Evaluasi

Evaluasi bertujuan untuk menilai hasil uji coba sistem yang dibuat apakah telah sesuai antara hasil sistem analisis sentimen dengan hasil sebenarnya. Confusion matrix adalah tabel yang digunakan untuk menganalisis seberapa baik keakuratan sebuah metode klasifikasi untuk memprediksi kelas suatu data.

III. METODOLOGI PENELITIAN.

A. Alur Penelitian

Dalam melakukan penelitian ini digunakan beberapa tahapan penelitian diantaranya adalah pengumpulan data, Proses Preprocessing, Pengimplementasian Metode TF-IDF untuk melakukan pembobotan kata, dan selanjutnya melakukan pengklasifikasian dengan menggunakan model Support Vector Machine. Dan terakhir melakukan pengujian Confusion Matrix dengan analisa Accuracy, Presisi, Recall dan F1-Score.



Gambar. 1 Gambaran umum alur penelitian

B. Pengambilan Data

Proses pengambilan data melibatkan pengumpulan data pendukung, yang diambil selama proses tersebut. Beberapa cara yang dapat digunakan adalah melalui studi literatur, dan salah satu cara yang dapat digunakan adalah dengan mengumpulkan dan mempelajari beberapa referensi dari berbagai sumber yang berkaitan dengan judul penelitian yang dilakukan. Dataset diambil melalui Twitter API dengan proses crawling. Dataset dalam bentuk format file csv. Ilustrasi pengambilan data melalui proses crawling dapat dilihat pada Gambar 2.

Id	Sentiment	Text	Tweet
0	1	0	Jelek filmnya... apalagi si ernest tidak bermu...
1	2	0	Film king Arthur ini film paling jelek dari se...
2	3	0	Sepanjang film aku berkata kasar terus pada ba...
3	4	0	aku tidak suka fast and furious..menurutku kok...
4	5	0	Aku tidak tahu film nya, kamu bilang perang pe...
...
195	196	1	Fargo juga adaptasi dari film yang cukup berha...
196	197	1	637.000 waw ini sangat keren film horor dengan...
197	198	1	@filmziarah film yang tenang dan menghanyutkan...
198	199	1	Film yang sangat menarik. Kisah cinta & keseti...
199	200	1	Nonton @filmziarah , film bagus, ada kali 5 me...

Gambar. 2 Hasil crawling data twitter

C. Pengolahan Data

1) Cleansing Data dan Case Folding

Didalam cleansing data terdapat proses case folding, dan hapus punctuation. Case folding sendiri merubah kata yang semula huruf kapital diubah menjadi lowercase agar seragam. Kemudian di dalam proses hapus punctuation terdapat fungsi penghapusan tanda baca, dan kode ASCII, hal ini dilakukan agar tanda baca tidak mengganggu proses pembobotan kata nanti. Berikut adalah gambar sebelum dan sesudah pemrosesan cleaning data.

Sentiment	Text	Tweet
0	0	Jelek filmnya... apalagi si ernest tidak bermu...
1	0	Film king Arthur ini film paling jelek dari se...
2	0	Sepanjang film aku berkata kasar terus pada ba...
3	0	aku tidak suka fast and furious..menurutku kok...
4	0	Aku tidak tahu film nya, kamu bilang perang pe...

Gambar 3. Sebelum Pemrosesan Cleansing Data

Sentiment	Text	Tweet
0	0	jelek filmnya apalagi si ernest tidak bermu...
1	0	film king arthur ini film paling jelek dari se...
2	0	sepanjang film aku berkata kasar terus pada ba...
3	0	aku tidak suka fast and furious menurutku kok...
4	0	aku tidak tahu film nya kamu bilang perang pe...

Gambar 4. Sesudah Pemrosesan Cleansing Data

2) Stopword

Menghapus stopword seperti di, ke, dari, serta, atau, berikan, kalau, akan, dan lain – lain. Berikut adalah gambar sebelum dan sesudah pemrosesan stopword data.

Sentiment	Text	Tweet
0	0	jelek filmnya apalagi si ernest tidak bermu...
1	0	film king arthur ini film paling jelek dari se...
2	0	sepanjang film aku berkata kasar terus pada ba...
3	0	aku tidak suka fast and furious menurutku kok...
4	0	aku tidak tahu film nya kamu bilang perang pe...

Gambar 5. Sebelum Pemrosesan Stopword

Sentiment	Text	Tweet
0	0	jelek filmnya si ernest bermutu sekali acti...
1	0	film king arthur film paling jelek seluruh cer...
2	0	sepanjang film aku berkata kasar terus bapaknya
3	0	aku suka fast and furious menurutku kok jelek...
4	0	aku tahu film nya kamu bilang perang perangan...

Gambar 6. Sesudah Pemrosesan Stopword

3) Stemming

Proses stemming ini mencari kata dasar dari sebuah kata. Seperti kata "menurut" setelah di stemming menjadi "turut". Berikut adalah gambar sebelum dan sesudah pemrosesan stemming data.

Sentiment	Text Tweet
0	jelek filmnya si ernest bermutu sekali acti...
1	film king arthur film paling jelek seluruh cer...
2	sepanjang film aku berkata kasar terus bapaknya
3	aku suka fast and furious menurutku kok jelek...
4	aku tahu film nya kamu bilang perang perang...

Gambar 7. Sebelum Pemrosesan Stemming

Sentiment	Text Tweet
0	jelek film si ernest mutu sekali actingnya fil...
1	film king arthur film paling jelek seluruh cer...
2	panjang film aku kata kasar terus bapak
3	aku suka fast and furious turut kok jelek itu ...
4	aku tahu film nya kamu bilang perang perang pe...

Gambar 8. Sesudah Pemrosesan Stemming

D. Proses Pembobotan Kata TF-IDF

Setelah melalui preprocessing, data berupa teks akan diubah menjadi angka dengan menghitung TF IDF. Gambar dibawah berikut adalah penghitungan TF-IDF. Sentimen negatif menunjukkan bahwa dokumen ke-3 menghasilkan nilai terbesar dengan nilai 1,431. Sedangkan untuk sentimen positif menghasilkan 2 dokumen yang mempunyai besar nilai yang sama yaitu 0,704.

Sentimen negatif

Kata kunci Suka Jelek
 Dokumen 1 (D1) jelek film si ernest
 Dokumen 2 (D2) king arthur film paling jelek seluruh cerita king arthur
 Dokumen 3 (D3) aku tidak suka Fast and Furious turut

Jumlah dokumen 3

Token	IK	DF	IDF	TF	DF	DF	IDF (log 6/014)	IK	DF	IDF	W	DF
jelek	1	1	3	0	2	1,5	0,176	0,4771	0,176	0,176	0	0
film	0	1	3	0	2	1,5	0,176	0	0,176	0,176	0	0
king arthur	0	1	3	0	1	3	0,4771	0	0,4771	0,4771	0	0
suka	1	0	0	1	3	0,4771	0,4771	0	0	0	0	0,4771
turut	0	0	0	1	3	0,4771	0	0	0	0	0	0,4771
fast and furious	0	0	0	0	3	0,4771	0	0	0	0	0	0,4771
Ranking	0,932	1,306	1,431									

Gambar 9. Perhitungan TF-IDF untuk Sentimen Negatif

Sentimen positif

Kata kunci Bagus Keren
 Dokumen 1 (D1) tonton film zarah film bagus kali 5 menit tonton tidak beranjak
 Dokumen 2 (D2) 637 000 war sangat keren film horor dengan jumlah tonton dalam waktu 4 hari
 Dokumen 3 (D3) turut saya bagus dan cukup keren dengan horor

Jumlah dokumen 3

Token	IK	DF	IDF	TF	DF	DF	IDF (log 6/014)	IK	DF	IDF	W	DF
tonton	0	1	3	0	2	1,5	0,176	0	0,176	0,176	0	0
film	0	2	3	0	2	1,5	0,176	0	0,352	0,352	0	0
bagus	1	1	3	0	2	1,5	0,176	0,176	0,176	0,176	0,000	0,176
keren	1	1	3	0	2	1,5	0,176	0,176	0	0	0	0,176
turut	0	0	0	1	3	0,4771	0	0	0	0	0	0,4771
horor	0	0	0	1	3	0,4771	0	0	0	0	0	0,4771
Ranking	0,704	0,704	1,000									

Gambar 10. Perhitungan TF-IDF untuk Sentimen Positif

Pada sentimen positif diatas terdapat nilai dari hasil perhitungan TF-IDF yang mempunyai hasil yang sama, maka kemudian untuk pengoptimalan dibantu oleh Algoritma Support Vector Machine (SVM) untuk mencari nilai terbesar dari sentiment positif. Gambar dibawah berikut menunjukkan hasil dari perhitungan algoritma SVM.

Hasil W ada yang sama maka kita menggunakan algoritma vector space model
 W kita pangkatkan 2

Lalu, jumlah kata di kali 0.031

HASIL AKHIR

Token	IK	DF	IDF	TF	DF	DF	IDF (log 6/014)	IK	DF	IDF	W	DF
tonton	0	1	3	0	2	1,5	0,176	0	0,176	0,176	0	0
film	0	2	3	0	2	1,5	0,176	0	0,352	0,352	0	0
bagus	1	1	3	0	2	1,5	0,176	0,176	0,176	0,176	0,000	0,176
keren	1	1	3	0	2	1,5	0,176	0,176	0	0	0	0,176
turut	0	0	0	1	3	0,4771	0	0	0	0	0	0,4771
horor	0	0	0	1	3	0,4771	0	0	0	0	0	0,4771
Ranking	0,932	1,306	1,431									

Gambar 11. Perhitungan Algoritma SVM

E. Evaluasi Data

Ada beberapa parameter yang bisa digunakan untuk mengevaluasi kinerja klasifikasi. Penelitian ini menggunakan Accuracy Precision, Recall, F1-Score sebagai parameter evaluasi untuk mengetahui seberapa baik sistem yang dibangun dan hasil yang diberikan oleh sistem.

$$Accuracy = (TP + TN) / (TP+FP+FN+TN)$$

$$Precision (P) = TP/(TP+FP)$$

$$Recall (R) = TP/(TP+FN)$$

$$F-1 Score = 2 * (R * P) / (R + P)$$

Evaluasi performa dilakukan pada hasil eksperimen sistem analisis sentimen dan pada hasil analisis sentimen responden. Hasil dari Precision, Recall, F1-Score menghasilkan nilai dari 0 sampai 1 dan hasil yang lebih baik adalah nilai yang mendekati 1[6].

IV. HASIL DAN PEMBAHASAN

Penelitian dilakukan menggunakan data yang diperoleh dari twitter berjumlah 200 kicauan, yang didapat melalui proses crawling. Data yang didapat berjumlah 200 dokumen, dengan mengambil kicauan dibatasi dari region Indonesia saja, maka didapat kicauan berbahasa Indonesia mengenai fenomena opini film di indonesia. Untuk proses analisa sentimen ini menggunakan bahasa pemrograman Python.

Data twitter kemudian diolah untuk dihilangkan simbol-simbol yang tidak berguna dan menentukan kata dasar. Berikut adalah tampilan tweet yang sudah melalui tahap preprocessing, akan disajikan dalam bentuk Word Cloud.



Gambar 12. Gambar Word Cloud untuk Sentimen Negatif



Gambar 13. Gambar Word Cloud untuk Sentimen Positif

Word Cloud akan menampilkan kata dengan frekuensi terbanyak, semakin banyak frekuensi kata tersebut semakin menonjol dibanding kata yang lain. Diketahui untuk word cloud yang warna merah kata yang sering muncul yaitu film, jelek dan kecewa ini bisa dikategorikan Negatif. Sedangkan untuk Word cloud warna biru kata yang sering muncul film, nonton, keren ini bisa dikategorikan positif.

Setelah tahu Word cloudnya selanjutnya masuk ke proses pembobotan kata menggunakan TF-IDF, kemudian setelah ketemu hasil TF-IDF nya baru diklasifikasikan menggunakan metode Super Vector Machine dan untuk evaluasi pengujian menggunakan confusion matrix, untuk hasilnya bisa dilihat pada tabel seperti dibawah ini.

TABEL I
TABEL HASIL PENGUJIAN

Accuracy	0.85
Precision	1.0
Recall	0.7
F1-Score	0.82

Berdasarkan Tabel 1 Metode Super Vector Machine dengan pembobotan TF-IDF menghasilkan nilai accuracy 0.85, nilai precision 1.0, nilai recall 0.7, dan nilai F1-Score 0.82. Untuk hasil ini sudah dikatakan baik tapi belum bisa dikatakan sempurna. Beberapa hal tersebut disebabkan oleh adanya suatu kata yang memiliki nilai sentimen positif jauh lebih besar dibandingkan dengan nilai sentimen negatif pada

dokumen yang kelas sebenarnya negatif, dan sebaliknya. Adanya kata-kata yang kurang baku di sebuah text tweet yang menyebabkan bermasalah pada saat proses pembobotan. Adanya kata-kata singkatan, sehingga banyak kata yang hilang setelah melewati proses Stemming

V. KESIMPULAN DAN SARAN

Sesuai dengan tujuan penelitian yaitu menganalisa sentimen opini film dengan menggunakan pembobotan TF-IDF dan algoritma svm telah berhasil dilakukan.

Dari hasil pengujian dan analisis dapat disimpulkan beberapa hal yaitu pengklasifikasian data opini film berbahasa indonesia berdasarkan sentimen dapat dilakukan dengan algoritma SVM. Semakin banyak data training yang digunakan maka akan mempengaruhi kinerja dari sistem. Hal ini dibuktikan dengan hasil nilai accuracy 0.85, nilai precision 1.0, nilai recall 0.7, dan nilai F1-Score 0.82. Hal ini sudah bisa dikatakan baik. Untuk penelitian kedepan disarankan adanya proses pengoptimalan sebuah tweet, agar kata tidak baku atau singkatan-singkatan bisa melewati proses preprocessing secara baik.

REFERENSI

- [1] B. Liu, *Sentiment Analysis: A Multi-Faceted Problem*. IEEE : Intelligent Systems, 2010.
- [2] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Science, 2011.
- [3] L. Francis and M. Flynn, *Text Mining Handbook, Casualty Actuarial Society E-Forum*. Spring, 2010.
- [4] I. Mathilda Yulietha and S. Al Faraby, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4740-4750, 2017.
- [5] M. Lestandy, A. Abdurrahim, and L. Syafa, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent," vol. 5, no. 10, pp. 802-808, 2021.
- [6] A. R. Dwi Pratiwi and E. Budi Setiawan, "Implementation of Rumor Detection on Twitter the SVM Classification Method," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 782-789, 2020, doi: <https://doi.org/10.29207/resti.v4i5.2031>.
- [7] M. W. Berry and J. Kogan, *Text Mining Application and Theory*. Wiley; 1st Edition, 2010.
- [8] R. Feldman and J. Sanger, *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press; 1st edition, 2006.
- [9] F. Tala, *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Institute for Logic, Language and Computation: Universiteit van Amsterdam, 2003.