

ANALISIS SENTIMEN FILM DI TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

¹⁾Mahdy Khairudin, ²⁾Ade Sukendar, ³⁾Asep Somantri

¹Teknik Informatika, Universitas Pasundan, ² Teknik Informatika, Universitas Pasundan,

³ Teknik Informatika, Universitas Pasundan

¹mr.mahdy5@gmail.com, ²ade.sukendar@unpas.ac.id, ³somantri@unpas.ac.id

ABSTRACT

Twitter is one of the social media that is often used to share information or so-called tweets in real time. Based on the results of research in 2013 by Semiocast, a social media research institute, said that the number of Twitter users in Indonesia was third after the United States and Japan with 6.5 percent of users and 1 billion tweets. Tweets can usually influence buying interest, voting for a political figure, or watching their movie. Tweets related to movies become a reference for other Twitter users if they want to watch the same movie. Because there is not service that provides movie references based on other people's comments that are relevant for Indonesian audiences. This study is to analyze the sentiment of tweets on Twitter in the case of Indonesian-language movies that produce movie ratings. The method used for sentiment analysis is the Support Vector Machine (SVM). The final result of this research is a model or SVM algorithm that can classify or analyze film sentiment based on tweets on Twitter by utilizing the Twitter API.

Keywords: *sentiment analysis, SVM, twitter API*

ABSTRAK

Twitter adalah salah satu media sosial yang sering digunakan untuk berbagi informasi atau biasa disebut tweet secara realtime. Berdasarkan hasil penelitian pada tahun 2013 oleh Semiocast, lembaga riset media sosial, mengatakan bahwa jumlah pengguna Twitter di Indonesia terbanyak ketiga setelah Amerika Serikat dan Jepang dengan angka 6,5 persen pengguna dan 1 miliar tweet. Tweet biasanya dapat mempengaruhi minat beli, memilih suatu tokoh politik, atau menonton film mereka. Tweet berkaitan dengan film menjadi referensi bagi pengguna Twitter yang lain apabila ingin menonton film yang sama. Karena tidak adanya layanan yang menyediakan referensi film berdasarkan komentar orang lain yang relevan untuk penonton Indonesia. Penelitian ini untuk melakukan analisis sentimen terhadap tweet di Twitter pada kasus film berbahasa Indonesia yang menghasilkan rating film. Metode yang digunakan untuk analisis sentimen adalah dengan Support Vector Machine (SVM). Hasil akhir dari penelitian ini adalah sebuah model atau algoritma SVM yang dapat mengklasifikasikan atau analisis sentimen film berdasarkan tweet di Twitter dengan memanfaatkan Twitter API.

Kata Kunci: *analisis sentimen, SVM, twitter API*

I. PENDAHULUAN

Twitter adalah salah satu media sosial yang sering digunakan untuk berbagi informasi atau biasa disebut tweet secara realtime. Berdasarkan hasil penelitian pada tahun 2013 oleh Semiocast, lembaga riset media sosial, mengatakan bahwa jumlah pengguna Twitter di Indonesia terbanyak ketiga setelah Amerika Serikat dan Jepang dengan angka 6,5 persen pengguna dan satu miliar tweet [1].

Pengguna Twitter yang semakin meningkat menimbulkan peningkatan tweet yang diposting.

Twitter juga adalah salah satu media sosial yang efektif dalam menampung opini dari mengenai produk, layanan termasuk film karena dalam memberikan komentar Twitter sangat cepat dan sentimen, evaluasi, tingkah laku dan emosi dari seseorang terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya [2]. Analisis sentimen berhubungan dengan pengelolaan data yang berkaitan dengan opini orang lain terhadap suatu entitas, telah dilakukan menggunakan teknik *machine learning* [3]. Dengan kata lain analisis sentimen adalah sub bidang

dari data mining [4]. Pada studi kasus film, tweet biasanya dapat mempengaruhi minat orang dalam menonton film tertentu khususnya yang menggunakan Bahasa Indonesia[5][6]. Tweet berkaitan dengan film menjadi referensi bagi pengguna Twitter yang lain apabila ingin menonton film yang sama.

II. METODE PENELITIAN

Berikut ini adalah langkah-langkah penelitian yang peneliti lakukan sebagai berikut.

1. Menentukan topik penelitian
Pada tahap ini dilakukan untuk menentukan topik penelitian yang akan diajukan oleh penulis.
2. Mengidentifikasi masalah
Pada tahap ini dilakukan untuk melakukan identifikasi masalah yang terjadi terkait dengan topik penelitian, serta solusi sementara yang akan diusulkan untuk mengatasi masalah tersebut.
3. Mengumpulkan jurnal dan makalah penelitian
Pada tahap ini dilakukan untuk mengumpulkan referensi yang relevan terkait masalah dan solusi sementara yang akan diusulkan.
4. Memahami teori yang terkait
Pada tahap ini dilakukan untuk memahami teori dan konsep yang terkait dengan masalah yang sudah diidentifikasi dan referensi yang didapatkan.
5. Mengeksplorasi tools terkait
Pada tahap ini dilakukan untuk mencari dan mengeksplorasi tools terkait dengan topik penelitian yang akan digunakan pada penelitian tersebut.
6. Mengumpulkan data penelitian
Pada tahap ini dilakukan untuk mengumpulkan data penelitian berupa data dari tweet berbahasa Indonesia di Twitter yang berkaitan dengan film dengan menggunakan Twitter API dan data training serta testing untuk membuat model analisis sentimen dari data penelitian ini [6].
7. Melakukan preprocessing
Pada tahap ini dilakukan preprocessing data agar data dari Twitter API dan data training serta data testing yang didapatkan menjadi terstruktur, bahasanya menjadi baku dan menghilangkan kesalahan penulisan. Dengan preprocessing ini membuat data dapat diolah oleh algoritma Support Vector Machine nya.

8. Melakukan analisis sentimen
Pada tahap ini dilakukan analisis sentimen data yang terdiri dari membuat model analisis sentimen dengan algoritma Support Vector Machine, menganalisis sentimen data film dari Twitter API dengan model yang telah dibuat dan pengitungan rating dari film.
9. Kesimpulan dan saran
Pada tahap ini dilakukan untuk memberikan kesimpulan dan saran dari hasil penelitian yang telah dilakukan.

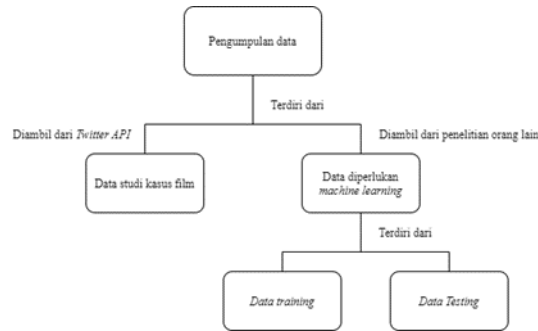


Gambar 1.
Metode Penelitian

III. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dapat diilustrasikan sebagai berikut:



Gambar 2. Ilustrasi Pengumpulan data

Pengumpulan data terdiri dari data yang diperlukan *machine learning* yang diambil dari penelitian orang lain. Data tersebut digunakan untuk *data training* dan *data testing*. Data kedua untuk studi kasus film yang diambil dari Twitter API [7].

3.2. Preprocessing Data

Teknik preprocessing data dapat mengubah data menjadi terstruktur sehingga dapat meningkatkan akurasi dari sebuah model atau algoritma machine learning SVM yang akan diimplementasikan [8]. Preprocessing data yang akan digunakan pada penelitian ini dapat digambarkan sebagai berikut [9]:



Gambar 3. Ilustrasi Preprocessing data

3.3. Analisis Sentimen

Menganalisis sentimen yang terdiri dari beberapa tahapan seperti pembuatan model SVM, pengklasifikasian sentimen dari data (positif dan negatif) dengan model SVM yang sudah dibuat dan penghitungan *rating* dari film [10]. Analisis sentimen terdiri dari tiga tahapan sebagai berikut

1. Pembuatan Model SVM

Tahapan pembuatan model atau algoritma SVM ini sebagai berikut

1.1 Membagi (*split*) data

Membagi (*split*) data dengan menggunakan *library* sklearn pada modul *model selection*. Hasil dari *split* data tersebut adalah Train_X, Test_X, Train_Y, Test_Y. Data x adalah data *tweet* sedangkan data y adalah label dari data *tweet* tersebut yaitu data sentimen (positif atau negatif).

1.2 Melakukan *label encoding*

Pada *machine learning* data seperti ini tidak dapat diproses karena kolom label nya yaitu sentimen masih berbentuk string atau termasuk *categorical data*, maka harus diubah menjadi data numerik dengan *label encoder* pada data Y yaitu Train_Y dan Test_Y. *Label encoder* adalah berfungsi untuk mengubah setiap nilai dalam kolom menjadi angka berurutan.

Contohnya:

Negative = 0

Postive = 1

Hasil dari *label encoding* sebagai berikut:

Id Sentiment		
0	1	0
1	2	0
2	3	0
3	4	0
4	5	0
...
195	196	1
196	197	1
197	198	1
198	199	1
199	200	1

Gambar 4. Hasil Label Encoder

1.3 Melakukan *feature extraction*

Komputer (mesin) tidak dapat mengolah data selain data numerik sehingga dibutuhkan suatu teknik untuk mengubah data teks menjadi numerik dengan *feature extraction*. *Feature extraction* suatu cara untuk mengekstrak data teks menjadi *feature* (ciri unik) yang merepresentasikan data aslinya.

Pada penelitian ini peneliti menggunakan TF-IDF karena sederhana namun relevan dalam melakukan pencocokan kata pada sebuah teks. *Term Frequency Inverse Document Frequency* (TF-IDF) ini merupakan metode untuk

memberikan bobot nilai suatu kata (*term*) pada suatu dokumen. Metode TF-IDF ini diterapkan pada data *tweet* menggunakan berdasarkan frekuensi data dan nilai dari suatu data yang digunakan.

1.4 Melatih Model SVM

Setelah mengubah data teks menjadi numerik maka melatih model atau algoritma *machine learning* menggunakan *Support Vector Machine* (SVM). SVM ini menggunakan *library* *sklearn* pada modul *svm* yaitu *SVC (Support Vector Classification)*.

Kemudian melatih model SVM tersebut dengan data *Train_X* yang sudah dilakukan *feature extraction* dan data *Train_Y* yang sudah dilakukan *label encoding*.

1.5 Mengevaluasi Model SVM

Setelah model SVM dilatih lalu model diuji dengan memprediksi klasifikasi data *Test_X* yang disesuaikan dengan jawaban sebenarnya yaitu *Test_Y*. Hal tersebut dilakukan untuk mengetahui apakah model SVM memiliki performa yang baik atau belum.

Untuk mengeceknya muncul istilah metrik yaitu suatu nilai yang merepresentasikan performa dari model yang dihasilkan. Terdapat 4 metrik yang umum digunakan yaitu *accuracy*, *precision*, *recall*, *f1 score*. Pada penelitian ini metrik yang digunakan yaitu *accuracy*, untuk menentukan presentase jumlah data yang diprediksi oleh model benar terhadap jumlah keseluruhan data. Berikut adalah hasil percobaan pertama dari prediksi model SVM.

Gambar 5. Metriks Prediksi Model SVM

Dari metrik diatas terdapat 4 data yang salah diprediksi oleh model yang nilai sebenarnya 1 atau positif sedangkan hasil prediksi nya 0 atau negatif. Ada 16 data yang benar diprediksi oleh model yang nilai sebenarnya 0 atau negatif. Terdapat juga 18 data yang benar diprediksi oleh model yang nilai sebenarnya 1 atau positif. Dan 2 data yang salah diprediksi oleh model yang nilai sebenarnya 0 atau negatif sedangkan hasil prediksi nya 1 atau positif.

Untuk akurasi dari model SVM pada percobaan pertama ini didapatkan akurasi sebesar 85% yang didapatkan dari jumlah data yang diprediksi benar dibagi dengan jumlah keseluruhan data *testing*.

1.6 Meningkatkan Akurasi Model

Untuk meningkatkan akurasi dari model dengan cara melakukan *tuning* pada *hyperparameter Support Vector Machine* (SVM).

Sebelum melakukan *tuning* pada *hyperparameter*, maka perlu mengetahui *hyperparameter* yang terbaik dengan menggunakan metode *grid search cross validation*. Cara kerja dari metode ini adalah dengan mengkombinasikan nilai pada *hyperparameter* dengan mencari *best score* terbaik.

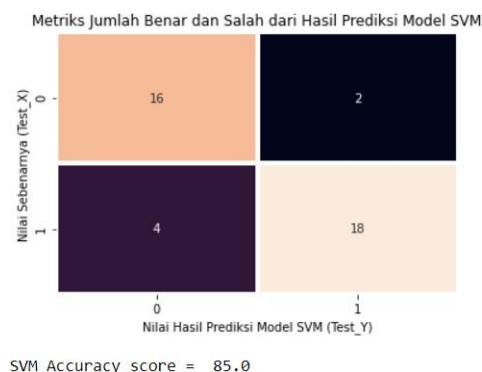
Pada metode ini juga dilakukan validasi silang pada data *testing* nya agar berubah dan mengacak data supaya menghindari *overfitting*.

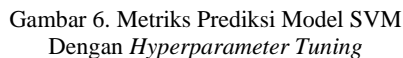
Berikut adalah hasil dari pengecekan *hyperparameter* terbaik dengan *grid search cross validation*:

Best C = 10
Best Kernel = rbf
Best Score = 0.85625

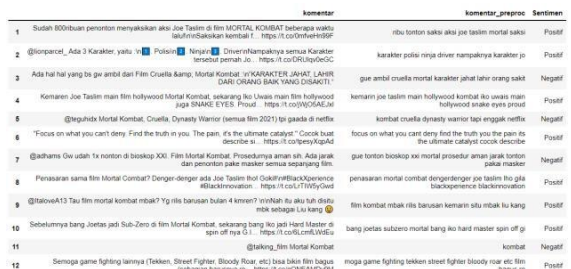
Gambar 10. Pengecekan *Hyperparameter* Terbaik

Kemudian memasukan *hyperparameter* terbaik pada model *Support Vector Machine* (SVM), maka didapatkan hasil Metrik Prediksi Model SVM sebagai berikut:





1. Pengklasifikasian Sentimen Data
Untuk pengklasifikasian sentimen data [11] menggunakan model SVM yang telah dibuat dan data yang digunakan adalah data dari Twitter API berbahasa Indonesia. Contoh data Twitter API yang digunakan adalah data film dengan kata kunci “Film Mortal Kombat” diakses pada 15 Juni 2021. Berikut ini adalah hasilnya:



Copyright@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 24-05-2023
Dipublikasikan pada 31-05-2023

Berikut ini adalah *prototype* aplikasi analisis sentimen film berbasis web pada halaman utama.



Rating Film Dengan Twitter API

Search Film: Search

Komentar Film black widow

Show 10 entries

No.	Tanggal	Komentar	Hasil Preprocessing	Sentimen
1	2018-12-28 12:23:18	i bismillah! Kidding abis banget ya? Udah banyak film kasetan. Black Widow, The Suicide Squad, Snake Eyes, Capet bukannya sampey lahman juga Shang-Chi, The Eternals, Spiderman sama Iron Man... View link	no handling buka sudah senak black widow the suicide squad snake eyes capet buka telah shangchi the eternal spiderman iron man	Negatif
2	2018-10-30 10:55:50	s/kar cinta Black Widow - 2021Ternyata, Natasha Romanoff alias "Black Widow" benar-mengalahkan gelag Hollywood saat sebuah komik yang berlatarnya dibikin dengan masa banyar... View link s/kar Cinta Cinta Kita, Film Neneklu Black Widow! View link View link	buka cinta black widow cinta natasha romanoff alias black widow telah no gelag hollywood benar benar komik cinta kita black widow	Negatif
3	2018-08-28 18:29:52	I @Bismillah kau terkejut udah 2017 ahk baik i juga ditonton RCTI SUMPIR KWOW, kau bisa maruk alah ih black widow no shang chi bang!	terkejut baik kau hayo-tonton rwa bangter sampai bangk film-maruk alah ih black widow sama shang chi bangter bang!	Positif

Berikut ini adalah *prototype* aplikasi analisis sentimen film berbasis web pada halaman hasil *preprocessing data*.

Pada pengujian ini peneliti akan melakukan percobaan dengan membandingkan kata kunci yang dimasukan. Pada percobaan ini dimasukan kata kunci

film yaitu “avengers” dan dibandingkan dengan kata kunci “avengers endgame” pada tanggal 25 agustus 2021. Hasil dari percobaan tersebut adalah sebagai berikut.

1. Film dengan kata kunci “avengers” memiliki:
 - a. Data komentarnya berjumlah 143 data.
 - b. Data sentimen positifnya berjumlah 43 data.
 - c. Data sentimen negatifnya berjumlah 100data.
 - d. *Rating* filmnya adalah 3.1.
2. Sedangkan jika kata kuncinya adalah “avengers endgame” memiliki:
 - a. Data komentarnya berjumlah 22 data.
 - b. Data sentimen positifnya berjumlah 11 data.
 - c. Data sentimen negatifnya berjumlah 11data.
 - d. *Rating* filmnya adalah 5.5.

IV. KESIMPULAN

Kesimpulan yang dapat diambil pada penelitian analisis sentimen film di *Twitter* yaitu tercapainya analisis sentimen film di *Twitter* menggunakan model atau algoritma SVM dengan akurasi model sekitar 87.5%. Data yang digunakan untuk training data sebanyak 160 data (80%) dan testing data sebanyak 40 data (20%).

Hyperparameter Terbaik yang digunakan untuk meningkatkan akurasi model yaitu C yang diisi 10 dan *kernel* diisi rbf yang membantu meningkatkan akurasi sebesar 2.5%. Dari pengujian tersebut dapat simpulkan bahwa jika kata kunci nya semakin spesifik pada satu film maka data komentar dari *Twitter* yang didapatkan lebih sedikit dibandingkan dengan kata kunci film secara umum

V. SARAN

Saran-saran untuk untuk penelitian lebih lanjut untuk menutup kekurangan penelitian.

Berikut ini merupakan saran yang dapat diberikan pada penelitian analisis sentimen film di *Twitter* yaitu *Twitter API* diharapkan menggunakan API yang dapat mengambil data *tweet* lebih dari hari ke belakang sehingga data yang didapatkan lebih banyak lagi. Berikutnya adalah pada saat melakukan proses klasifikasi diharapkan memisahkan *tweet* yang mengandung opini dan tidak sehingga proses klasifikasi akan lebih optimal serta meningkatkan akurasi dalam menghitung *rating* film. Selain itu

diharapkan juga menggunakan *hyperparameter* yang lainnya seperti gamma dan lainnya.

DAFTAR PUSTAKA

- [1] Rindu P. Hestya, “Indonesia Pengguna Twitter Nomor 3 di Dunia”, 2013. [Online]. Available: <https://tekno.tempo.co/read/538043/indonesi-a-pengguna-twitter-nomor-3-di-dunia> [Accessed: 18-Ags-2020].
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. 2012.
- [3] Akshay Amolik, et al., 2016. “Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques,” *International Journal of Engineering and Technology (IJET)*, Vol. 7, No. 6, 2016.
- [4] D. Hand, H. Mannila, & P. Smyth, *Principles of Data Mining*. A Bradford Book The MIT Press, 2001.
- [5] Eyheramendy, S., Lewis, D. D., and Madigan, D, “On the Naive Bayes Model for Text Categorization”. Ninth International Workshop on Artificial Intelligence and Statistics. 2013.
- [6] Hadna, N. et al., 2016. “Studi Literatur Tentang Perbandingan Metode Proses Analisis Sentimen di Twitter”. Fakultas Teknik, Universitas Gadjah Mada, 2016.
- [7] Twitter, “Twitter API Docs”, [Online]. Available: <https://developer.twitter.com/en/docs> [Accessed: 20-Mar-2021].
- [8] Jiawei Han, M. Kamber, Jian Pei, *Data Mining Concepts and Techniques Third Edition*, Morgan Kaufmann Publisher, 2012.
- [9] Prananda A., Rizal S. P., M. Ali Fauzi, “Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku”, Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, 2017.
- [10] Nugroho, Witarto, & Handoko, “Application of Support Vector Machine in Bioinformatics”, *Proceeding of Indonesian Scientific Meeting in Central*. 2003.
- [11] Nuvirta M., Lukito E. N., Bimo S. H., “Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating”, Jurusan Teknologi Informasi, Universitas Gajah Mada, 2014.