

ANALISIS SENTIMEN PADA *TWITTER* MENGENAI NETFLIX DIBLOKIR TELKOM MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Rizqa El Husna¹, Rochdi Wasono², M. Al Haris³
Universitas Muhammadiyah Semarang

rizqaelhusna82@gmail.com

Abstrak. Netflix merupakan salah satu media *streaming online* yang beroperasi di Indonesia sejak 2016. Biaya langganan Netflix yang mahal dibandingkan media *streaming* lain tidak menghalangi pertumbuhan Netflix di Indonesia yang semakin meningkat setiap tahunnya. Pelanggan PT. Telkom Indonesia tidak bisa mengakses Netflix karena diblokir oleh PT. Telkom Indonesia. Hal ini mengakibatkan pelanggan Telkom tidak bisa mengakses Netflix menggunakan provider Telkomsel dan Indihome. *Twitter* merupakan media sosial yang memungkinkan penggunanya membaca, meneruskan, dan mengirim pesan berbasis teks. Pada media sosial *Twitter* banyak pelanggan Telkom yang menyampaikan opini mengenai Netflix diblokir Telkom. Berdasarkan opini publik di media sosial *Twitter* metode klasifikasi cocok digunakan untuk mengetahui sentimen publik. Metode *Support Vector Machine (SVM)* sering digunakan pada analisis klasifikasi sentimen. SVM merupakan usaha mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah kelas dengan memaksimalkan jarak antar kelas. Tujuan dari penelitian ini adalah mengetahui tingkat akurasi SVM dan mengetahui tampilan *word cloud* berdasarkan masing-masing sentimen. Jumlah *tweet* yang diteliti pada penelitian ini sebanyak 1771 *tweet*. Perbandingan data *training* dan data *testing* pada penelitian adalah 70% sebanyak 1241 *tweet* dan 30% sebanyak 530 *tweet*. Klasifikasi sentimen metode *Support Vector Machine* menggunakan *Kernel Radial Basic Function (RBF)* diperoleh akurasi sebesar 85.92%. Pada sentimen positif kata “buka”, “tidak”, “indihome”, merupakan kata yang paling sering digunakan. Sedangkan pada sentimen negatif kata “buka”, “pakai”, “telkom” merupakan kata yang paling sering digunakan

Kata Kunci: *Netflix, Twitter, Analisis Sentimen, Support Vector Machine (SVM).*

1. Pendahuluan

Menonton merupakan salah satu hiburan yang sangat digemari oleh masyarakat. Di zaman digital yang serba cepat dan serba ada, banyak masyarakat yang beralih dari menonton televisi ke menonton di layanan *digital streaming*. Pesatnya perkembangan internet di dunia, menghadirkan banyak usaha dan jasa yang mulai bergerak dari *offline* ke *online*, salah satunya adalah layanan media *streaming*. Menurut *We Are Social* 64% penduduk Indonesia telah terkoneksi dengan internet dari total populasi di Indonesia yang mencapai 272 juta jiwa. Semakin pesatnya pengguna internet, semakin berkembang pula pengguna media sosial. Hal ini dibuktikan dengan 59% atau 160 juta penduduk Indonesia aktif menggunakan media sosial.



Gambar 1. Jumlah Pengguna Internet di Indonesia Tahun 2020

Media sosial merupakan media *online*, dimana penggunanya bisa mengakses informasi, berbagi, berpartisipasi, dan menciptakan konten media sosialnya. Salah satu media sosial yang banyak digunakan adalah *Twitter*. *Twitter* adalah media sosial yang penggunanya bisa mengekspresikan tulisannya dalam 140 karakter. Hal ini membuat *Twitter* salah satu wadah bagi masyarakat untuk menyampaikan opininya. Pada tahun 2020 jumlah pengguna aktif harian *Twitter* kini mencapai 166 juta, meningkat 24% dari 134 juta pada 2019 lalu. Salah satu penyedia layanan media *streaming* untuk menonton *online* adalah Netflix. Pada tahun 2016 Netflix hadir di Indonesia, kehadiran Netflix merupakan berkah bagi para pecinta film. Netflix adalah layanan yang memungkinkan pengguna menonton tayangan kesukaan dimanapun, kapanpun, dan lewat medium apapun (*smartphone*, *smart televisi*, *tablet*, *Personal Computer*, dan laptop). Sejak beroperasi di Indonesia pada tahun 2016 sampai Januari 2020 pengguna Netflix sudah mencapai

900 juta pelanggan. Pertambahan pelanggan Netflix dalam satu tahun terakhir mencapai 80% (dapat dilihat pada gambar 2). Pesatnya pertumbuhan pelanggan Netflix di Indonesia memicu perusahaan layanan streaming lainnya untuk melebarkan bisnisnya di Indonesia, diantaranya iFlix, Viu, HOOQ, Goplay.



Gambar 2. Jumlah Pelanggan *Streaming* Netflix di Indonesia

Banyaknya layanan *streaming digital* yang ada di Indonesia seperti iFlix, HOOQ, Viu, dan Goplay memberikan pilihan yang bervariasi bagi para pecinta film. Biaya langganan Netflix lebih mahal dibandingkan layanan *streaming* lainnya tidak menghalangi pertumbuhan Netflix di Indonesia yang semakin meningkat setiap tahunnya. Pada penghujung tahun 2019 Netflix ramai diperbincangkan oleh pengguna *Twitter* karena pelanggan Telkom group tidak bisa mengakses Netflix. Hal ini menyebabkan para pengguna *Twitter* kembali mengeluarkan opini perihal Netflix yang diblokir oleh Telkom.

Berdasarkan latar belakang di atas perlu dilakukan penelitian untuk mengetahui tanggapan masyarakat berdasarkan sentimen di media sosial (*Twitter*) terhadap permasalahan Netflix yang diblokir oleh Telkom group. Analisis yang cocok digunakan untuk mengetahui sentimen masyarakat terhadap suatu permasalahan, produk, atau jasa adalah analisis sentimen. Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks, sentimen biasanya digunakan untuk mengetahui opini masyarakat terhadap produk, layanan, politik, atau topik lainnya. Metode klasifikasi dalam ilmu statistika yang sering digunakan pada analisis sentimen diantaranya adalah *Naïve Bayes*, *Support Vector Machine*, *K Nearest Neighbour*, Asosiasi, dll.

Metode *Support Vector Machine (SVM)* sering digunakan karena sangat cepat dan akurat pada klasifikasi data (Feldman & Sanger, 2007). Konsep SVM merupakan usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas sentimen. SVM berusaha menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak antar kelas. Pada penelitian Ditia Yosmita Praptiwi (2018) Analisis sentimen *online review* pengguna *e-commerce* menggunakan metode *Support Vector Machine* dan *Maximum Entropy*, nilai akurasi metode *Support Vector Machine* menggunakan kernel linear, Polynomial, RBF, Sigmoid adalah 90,39%, 55,32%, 91,95%, dan 91,17%. Berdasarkan penelitian tersebut kernel *Radial Basis Function (RBF)* lebih tinggi dibandingkan kernel lainnya yakni sebesar 91,95%. Pada penelitian ini, data yang digunakan adalah variabel independen yang diperoleh dari kata *tweet*. Kemudian dilakukan *praprocess text (cleansing, case folding, tokenizing, filtering, dan stemming)* dari kata *tweet* untuk mengklasifikasi sentimen *tweet* (positif dan negatif). Penelitian ini bertujuan untuk mengetahui kinerja metode *Support Vector Machine* dalam mengklasifikasikan sentimen pengguna *Twitter* terhadap aplikasi Netflix yang diblokir oleh Telkom group dan mengetahui visualisasi *word cloud* masing-masing sentimen.

2. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah seluruh *tweet* dengan *keyword* “Netflix blokir” di *Twitter* dari tanggal 1 Januari 2020 sampai dengan 30 April 2020 sebanyak 1771 *tweet*. *Scrapping* data menggunakan *software* Python. Pada penelitian ini data dibagi menjadi data *training* dan data *testing* dengan perbandingan 70% data *training* sebanyak 1241 *tweet* dan 30% data *testing* sebanyak 530 *tweet*. Perbandingan 70% data *training* dan 30% data *testing* dipilih berdasarkan nilai akurasi tertinggi dibandingkan perbandingan lainnya.

Tweet yang diperoleh, diberi label positif dan negatif. Data yang diperoleh dibagi menjadi data *training* (data latih) dan data *testing* (data uji). Data *training* digunakan untuk melatih algoritma, sedangkan data *testing* digunakan untuk menguji performa dan kebenaran dalam mengklasifikasikan sentimen. Struktur data yang digunakan dalam penelitian ini adalah *tweet* yang sudah dilakukan *text*

preprocessing yang terdiri dari variabel prediktor yaitu kata dasar setiap *tweet* dan variabel respon yaitu klasifikasi sentimen *tweet* (positif dan negatif).

Tahapan analisis yang digunakan dalam penelitian ini adalah :

- a. *Scrapping* data *tweet* dengan keyword “Netflix blokir”
- b. Melakukan tahap *text preprocessing*, yaitu : *cleaning*, *case folding*, *tokenizing*, *stopwords*, dan *stemming*
- c. Melakukan pelabelan sentimen pada *tweet*
- d. Membagi data *testing* dan data *training*
- e. Klasifikasi data menggunakan metode *Support Vector Machine kernel RBF*
- f. Melakukan pengujian performa metode NBC dan SVM berdasarkan tingkat *accuracy*, *precision*, *recall*, dan *F Measure*
- g. Melakukan visualisasi *tweet* dengan *word cloud*
- h. Interpretasi hasil analisis dan menarik kesimpulan

3. Hasil dan Pembahasan

3.1 Text Preprocessing

Tabel 1. Tahapan *Text Preprocessing*

<i>Text Preprocessing</i>	<i>Tweet</i>
<i>Cleaning</i>	@IndiHome oh indihome!! Kapan dibuka blokir netflix nya
<i>Case Folding</i>	oh indihome!! kapan dibuka blokir netflix nya
<i>Tokenizing</i>	oh indihome!! kapan dibuka blokir netflix nya
<i>Filtering</i>	oh indihome terus kapan dibuka blokir netflix nya
<i>Stemming</i>	indihome kapan buka blokir netflix nya

Berdasarkan tabel di atas untuk huruf yang berwarna merah adalah huruf yang mengalami *text processing*.

3.2 Pelabelan Data

Langkah selanjutnya adalah pelabelan pada kelas sentimen menggunakan kategori positif dan kategori negatif. Tanggapan tanggapan positif pengguna Twitter mengenai Netflix yang diblokir sebanyak 1.114 tweets sedangkan sentimen negatif sebanyak 657 tweets.

Tabel 2. Pelabelan *Tweet*

Sentimen	Skor	Tanggapan / <i>Tweet</i>
Positif	1	indihome kapan buka blokir netflix nya
Negatif	-1	kemendikbud tak ikut campur soal telkom blokir Netflix

Contoh perhitungan kelas sentimen berdasarkan *tweet* “kemendikbud tak ikut campur soal telkom blokir netflix”, terdapat 1 kata negatif dan 0 kata positif yang terdeteksi yaitu “tak” sebagai kata negatif.. Adapun rumus perhitungan skor sentimen yang digunakan dalam proses pelabelan adalah sebagai berikut:

Skor = (Jumlah kata positif) – (Jumlah kata negatif). Data teks terlebih dahulu akan diubah ke dalam bentuk vector menggunakan teknik.

3.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Simulasi TF-IDF akan dijelaskan perhitungan bobot dokumen terhadap *query* sesuai dengan rumus yang telah dijelaskan. Contoh *query terms* (Q) yang digunakan adalah “buka”, “tak”. Misal terdapat dua buah dokumen yaitu:

Dokumen 1 (d_1) = indihome kapan buka blokir netflix nya

Dokumen 2 (d_2) = kemendikbud tak ikut campur soal telkom blokir netflix

Adapun beberapa *term* (dokumen *term*) yang didapatkan dari kedua dokumen tersebut yaitu:

- a. Jumlah kemunculan kata ”buka” dalam d_1 adalah tidak ada sehingga $tf_j = tf_{buka} = 1$
- b. Total seluruh dokumen yang ada yaitu sebanyak tiga dokumen sehingga $D=2$
- c. Dari ketiga dokumen tersebut, terdapat satu dokumen yang memuat kata “buka” sehingga $df_j = df_{buka} = 1$

Tabel 3. Simulasi Perhitungan TF-IDF

Q	Tf		df	IDF	W	
	d_1	d_2			d_1	d_2
buka	1	0	1	0.301	0.301	
tak	0	1	1	0.301		0.301
Nilai Bobot Masing-masing Dokumen					$w(d_1)$	$w(d_2)$
					0.301	0.301

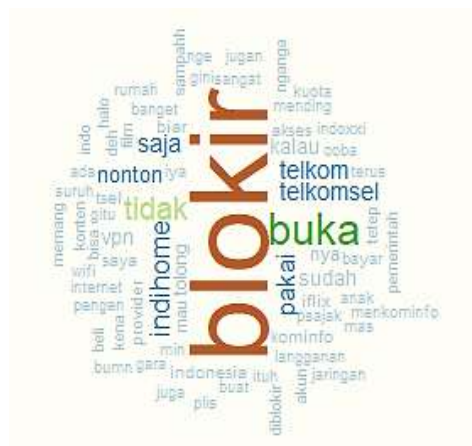
3.4. Support Vector Machine

Langkah awal dalam *Support Vector Machine* adalah membagi data *training* dan data *testing*. Kemudian merubah teks menjadi vektor dan pembobotan kata dengan menggunakan TF-IDF.

Aktual	Prediksi		Jumlah
	Positif	Negatif	
Positif	318	26	344
Negatif	48	138	186
<i>Accurasy</i>	86.04 %		530

Berdasarkan hasil perhitungan menggunakan *confusion matrix* pada algoritma *Support Vector Machine Kernel RBF* data *tweets* diatas prediksi benar sebanyak 456 dan prediksi salah sebanyak 74. Hasil klasifikasi dengan algoritma *SVM* untuk Netflix blokir diperoleh nilai *accuracy* sebesar 86.04%. *Precession* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precession* dengan menggunakan metode *SVM* sebesar 86.49%. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* menggunakan metode *SVM* diperoleh sebesar 92.44%. *F Measure* merupakan perbandingan rata-rata presisi dan recall yang dibobotkan, nilai *F Measure* pada metode *SVM* sebesar 89.36%.

3.5 Visualisasi Word Cloud



Gambar 3. *Word Cloud* Sentimen Positif Negatif



Gambar 4. *Word Cloud* Sentimen Negatif

Kata blokir yang berukuran paling besar dibandingkan dengan kata yang lain, menunjukkan bahwa kata “blokir” yang paling banyak digunakan oleh pengguna *Twitter*. Hal ini dikarenakan “blokir” merupakan kata keyword pada saat *scrapping* data. Pada *word cloud* sebelah kiri yaitu sentimen positif *term* “buka” dan *term* “tidak” merupakan kata yang paling banyak muncul setelah blokir. Hal ini menunjukkan bahwa pada sentimen positif pengguna *Twitter* menginginkan Telkom membuka blokir Netflix. Selain itu kata “tidak” menunjukkan pengguna *Twitter* tidak setuju dengan Netflix yang diblokir.

Pada *word cloud* negatif kata “buka” dan kata “pakai” merupakan dua kata yang paling banyak muncul dibandingkan kata yang lainnya. Kata “buka” banyak muncul pada sentimen negatif dikarenakan penentuan klasifikasi berdasarkan kata dalam suatu kalimat. Hal ini sudah dijelaskan pada tahap pelabelan kata. Selanjutnya kata “pakai” merupakan kata kedua yang banyak muncul. Hal ini merujuk pada tanggapan pengguna *Twitter* yang banyak memakai vpn untuk membuka Netflix bagi pelanggan Telkom. Pada sentimen negatif menunjukkan pengguna *Twitter* setuju dengan kebijakan Telkom yang memblokir Netflix bagi pelanggannya.

4. Kesimpulan

- a. Hasil ketepatan klasifikasi menggunakan *Support Vector Machine* dalam mengklasifikasikan sentimen pengguna *Twitter* mengenai Netflix yang diblokir diperoleh akurasi sebesar 86.04%
- b. Berdasarkan hasil tampilan *word cloud* pada sentimen positif kata “buka”, “tidak”, “indihome”, “pakai”, “telkomsel”, “telkom”, “saja”, “nonton” merupakan kata yang paling sering digunakan. Sedangkan pada sentiment negatif kata “buka”, “pakai”, “telkom” merupakan kata yang paling sering digunakan.

5. Ucapan Terima Kasih

Pada kesempatan ini penulis ingin menyampaikan terima kasih dan penghargaan yang setinggi-tingginya Bapak Dr. Rochdi Wasono, M.Si., dan Bapak M. Al Haris, S.Si., M.Si., selaku Dosen Pembimbing yang telah meluangkan waktu,

tenaga dan pikirannya untuk memberikan bimbingan dan pengarahan dalam penyusunan jurnal ini. Terima kasih kepada semua pihak yang telah membantu penyelesaian jurnal ini.

Daftar Pustaka

- Berry, M.W. & Kogan, J. 2010. *Text Mining Application and theory*. WILEY : United Kingdom.
- Boy, Utomo Manalu. 2014. *Analisis Sentimen Pada Twitter Menggunakan Text Mining*. Skripsi. Program Studi Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara
- Fikria, Norma,. 2018. *Analisis Klasifikasi Sentimen Review Aplikasi E-ticketing Menggunakan Metode Support Vector Machine dan Asosiasi*, Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia Yogyakarta.
- Liu, Bing. 2012. *Sentimen Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- Praptiwi, Ditia Yosmita. 2018. *Analisis Sentimen Online Review Pengguna E-commerce menggunakan metode Support Vector Machine dan Maximum Entropy*. Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia Yogyakarta.
- Siang, J. J. (2005). *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB*. Yogyakarta: ANDI.
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung:Informatika Bandung
- Ulwan, M. N. 2016. *Pattern Recognition pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association*. Skripsi: Program Studi Statistika Universitas Islam Indonesia Yogyakarta
- Wearesocial. 2020. *Digital in 2020*. <https://wearesocial.com/digital-2020>. Diakses pada tanggal 30 Juni 2020