



**ANALISIS KLASIFIKASI SENTIMEN PADA *TWITTER* MENGENAI
NETFLIX YANG DIBLOKIR OLEH TELKOM MENGGUNAKAN
NAÏVE BAYES CLASSIFIER DAN *SUPPORT VECTOR MACHINE***

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

RIZQA EL HUSNA

B2A219030

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS MUHAMMADIYAH SEMARANG

2020

PERSETUJUAN PEMBIMBING

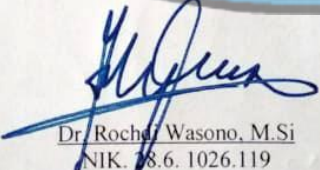
Nama : Rizqa El Husna
NIM : B2A219030
Program Studi : SI-Statistika
Judul Skripsi : Analisis Klasifikasi Sentimen Pada *Twitter* Mengenai
Netflix yang Diblokir oleh Telkom Menggunakan *Naïve*
Bayes Classifier dan *Support Vector Machine*

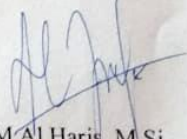
Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia
ujian Skripsi Program Sarjana Statistika

Semarang, September 2020

Pembimbing Utama

Pembimbing Pendamping


Dr. Rochet Wasono, M.Si
NIK. 88.6. 1026.119


M. Al Haris, M.Si
NIK. CP. 1026.070

**SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH**

Yang bertanda tangan di bawah ini, saya :

Nama : Rizqa El Husna
NIM : B2A219030
Fakultas/Jurusan : FMIPA/S1 Statistika
Jenis Penelitian : Skripsi
Judul : Analisis Klasifikasi Sentimen Pada *Twitter* Mengenai Netflix yang Diblokir oleh Telkom Menggunakan *Naïve Bayes Classifier* dan *Support Vector Machine*
Email : rizqaelhusna82@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada Perpustakaan UNIMUS atas penulisan karya ilmiah saya, demi perkembangan ilmu pengetahuan
2. Memberikan hak menyimpan, mengalih mediakan/mengasah formatkan, mengelola dalam bentuk pangkalan data (database), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada Perpustakaan UNIMUS, tanpa perlu izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak Perpustakaan UNIMUS dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 27 September 2020

Yang membuat pernyataan,


Rizqa El Husna
(B2A219030)



ANALISIS KLASIFIKASI SENTIMEN PADA *TWITTER* MENGENAI NETFLIX YANG DIBLOKIR OLEH TELKOM MENGGUNAKAN *NAÏVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE*

Rizqa El Husna¹, Rochdi Wasono, M. Al Haris³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : 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. PT. Telkom Indonesia yang memblokir Netflix bagi pelanggannya, mengakibatkan pelanggan Telkom tidak bisa mengakses Netflix menggunakan provider Telkomsel dan Indihome. *Twitter* merupakan media sosial yang memungkinkan penggunaanya membaca, meneruskan, dan mengirim pesan berbasis teks. Pada media sosial *Twitter* banyak pelanggan Telkom yang menyampaikan opini mengenai perihal Netflix diblokir. Berdasarkan opini publik di media sosial *Twitter* metode klasifikasi cocok digunakan untuk mengetahui sentimen publik. Metode *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)* sering digunakan pada analisis sentimen. NBC merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data *testing* pada kategori yang paling tepat. SVM merupakan usaha mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah kelas dengan memaksimalkan jarak antar kelas. Jumlah *tweet* yang diteliti pada penelitian ini sebanyak 1789 *tweet*. Perbandingan data *training* dan data *testing* pada penelitian adalah 70% dan 30%. Pada pengujian data *testing* sebanyak 530 data menggunakan *Naïve Bayes Classifier* akurasi klasifikasi yang didapat yaitu sebesar 75.47%. Sedangkan metode *Support Vector Machine* menggunakan *Kernel Radial Basic Function (RBF)* diperoleh akurasi sebesar 85.92%. Berdasarkan nilai akurasi, metode *Support Vector Machine* menggunakan *Kernel RBF* memiliki performa lebih baik dalam mengklasifikasi data sentimen.

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

ABSTRACT

Netflix is one of the online streaming media, has been operated in Indonesia since 2016. The cost of a Netflix subscription is expensive compared to other streaming media. That does not hinder the growth of Netflix in Indonesia. PT. Telkom Indonesia blocks Netflix for its customers, has resulted in Telkom subscribers being unable to access Netflix using Telkomsel and Indihome providers. Twitter is a social media that allows its users to read, forward and send text-based messages. On social media Twitter, many Telkom's customers expressed their opinions regarding Netflix being

blocked. Based on public opinion on Twitter, the classification method is suitable to determine public sentiment. The Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) methods are often used in sentiment analysis. NBC is an algorithm that is used to find the highest probability value for classifying testing data into the most appropriate category. SVM is an effort to find the best hyperlane that functions as a separator of two classes by maximizing the distance between classes. The number of tweets studied in this research were 1789 tweets. The comparison of training data and testing data in this research was 70% and 30%. Ratio the testing data as many as 530 data using the Naïve Bayes Classifier, the classification accuracy obtained is 75.47%. While the Support Vector Machine method using the Kernel Radial Basic Function (RBF) obtained an accuracy of 86.04%. Based on the accuracy value, the Support Vector Machine method using the RBF Kernel has a better performance in classifying sentiment data.

Keywords: Netflix, Twitter, Sentimen Analysis, Support Vector Machine (SVM), Naive Bayes Classifier (NBC)

PENDAHULUAN

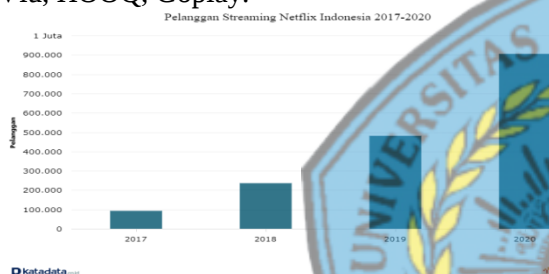
Menonton merupakan salah satu hiburan yang sangat digemari oleh masyarakat. Di zaman digital yang serba cepat dan serba ada, banyak masyarakat yang beralih dari menonton televisi ke menonton di layanan *digital streaming*. Pesatnya perkembangan internet di dunia, menghadirkan banyak usaha dan jasa yang mulai bergerak dari *offline* ke *online*, salah satunya adalah layanan media *streaming*.

Berdasarkan informasi yang dipublikasi oleh website *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.



Media sosial merupakan media *online*, dimana penggunaanya bisa mengakses informasi, berbagi, berpartisipasi, dan menciptakan konten media sosialnya. Salah satu media sosial yang banyak digunakan adalah *Twitter*. *Twitter* adalah media sosial yang penggunaanya 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 di bawah). Pesatnya pertumbuhan pelanggan Netflix di Indonesia memicu perusahaan layanan streaming lainnya untuk melebarkan bisnisnya di Indonesia, diantaranya iFlix, Viu, HOOQ, Goplay.



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. Model layanan Netflix yang perlu disesuaikan Undang-Undang (UU) Informasi dan Transaksi Elektronik, UU Penyiaran, UU Pornografi, UU Perlindungan Anak. Ternyata Netflix juga diburu oleh Kementerian Keuangan karena sejak beroperasi di Indonesia dari tahun 2016, Netflix belum menyetorkan pajak kepada Kementerian Keuangan. Pada bulan Januari

2020 Netflix kembali menjadi *trending topic* di Indonesia karena bekerjasama dengan Kementerian Pendidikan dan Kebudayaan (Kemendikbud) untuk menumbuhkan perfilman Indonesia. Keputusan Kemendikbud menjadi sorotan karena Netflix yang masih diburu oleh Kementerian Keuangan, diblokir oleh Telkom, dan diawasi oleh KPI (Komisi Penyiaran Indonesia) dalam hal sensor film yang akan tayang di Netflix. Hal ini menyebabkan “#Netflix Tidak Aman” kembali menjadi *trending topic* di Indonesia dengan jumlah *tweet* sebanyak 14,4 ribu *tweet*. Banyaknya masyarakat yang *tweet* dengan “#Netflix Tidak Aman”, membuktikan tingginya respon masyarakat terhadap Netflix. Sehingga 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 *Naïve Bayes Classifier* (NBC) banyak digunakan pada penelitian *Text Mining* karena algoritmanya yang sederhana, sehingga mudah dipahami dan memiliki akurasi yang tinggi (Rish, 2006). Sedangkan metode *Support Vector Machine* sering digunakan karena sangat cepat dan akurat

pada klasifikasi data (Feldman & Sanger, 2007). 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 Nur Khotimah (2019) *Sentimen Analysis of E-Commerce Brand Review Using Multinomial Text Naïve Bayes* tingkat akurasi menggunakan metode *Naïve Bayes Multinomial* mencapai 91.95%.

Pada penelitian ini, data yang digunakan adalah variabel independen yang diperoleh dari kata dasar *tweet*. Kemudian dilakukan *praprocess text* dari kata *tweet* untuk mengklasifikasi sentimen *tweet* (positif dan negatif). Penelitian ini bertujuan untuk mengetahui analisis sentimen masyarakat dari media sosial *Twitter* terhadap aplikasi *Netflix* yang diblokir oleh Telkom group. Metode klasifikasi yang digunakan pada penelitian ini adalah *Naïve Bayes Classifier* dan *Support Vector Machine* menggunakan kernel *Radial Basis Function (RBF)*.

TINJAUAN PUSTAKA

1. Netflix

Netflix adalah layanan *streaming* yang menawarkan berbagai acara televisi, film, film dokumenter, dan anime yang diakses dengan perangkat yang terhubung ke internet. Pengguna Netflix bisa menonton sepuasnya, kapan pun, dimana pun, dengan media apa pun dengan biaya langganan per bulan

2. Twitter

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunaanya untuk mengirim dan

membaca pesan berbasis teks dengan batas 280 karakter.

3. Analisis Sentimen

Analisis sentimen adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif

4. Text mining

Text mining adalah sekumpulan bahasa alami yang tidak terstruktur berupa berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, dll Pada *text mining* dilakukan tahapan *text preprocessing* yaitu

- Case Folding* adalah suatu bentuk *text preprocessing* yang paling sederhana dan efektif meskipun sering diabaikan.
- Tokenizing* adalah proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token.
- Filtering* adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).
- Stemming* adalah proses mendapatkan kata dasar dengan menghilangkan imbuhan kata.

5. TF-IDF

Tujuan dari TF-IDF adalah untuk menemukan jumlah kata yang diketahui (tf) setelah dikalikan dengan beberapa banyak *tweet* dimana suatu kata tersebut

muncul (idf). Berikut rumus menentukan pembobot dengan TF-IDF

$$idf = \log \frac{N}{df_j}$$

$$w_{ij} = tf_{ij} \times idf_j$$

6. Naïve Bayes Classifier (NBC)

Naive Bayes Classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam algoritma NBC setiap dokumen direpresentasikan dengan pasangan atribut “ $a_1, a_2, \dots, a_n$ ” dimana a_1 adalah kata pertama, a_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori tweet.

Nilai (V_j) dihitung pada saat *training*, didapat dengan rumus sebagai berikut :

$$P(V_j) = \frac{|doc_j|}{|training|}$$

Untuk setiap probabilitas kata a_i untuk setiap kategori $P(a_i|v_j)$, dihitung pada saat *training*.

$$P(a_i|v_j) = \frac{n_i + 1}{|n + kosakata|}$$

Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}).

$$V_{MAP} = \arg \max_i P(V_j) P(a_i|v_j)$$

7. Support Vector Machine (SVM)

SVM adalah usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space. SVM berusaha menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak antar kelas SVM. Pada data non linear menggunakan salah satu metode Kernel. Pada

penelitian ini penulis menggunakan metode Kernel RBF

$$K(x, y) = \exp \frac{-(2.1) \|x - y\|^2}{2 \cdot \sigma^2}$$

8. Pengukuran Performa

Pengukuran performa dilakukan untuk menentukan bagus/tidaknya sebuah klasifikasi. Terdapat beberapa cara untuk mengukur performa diantaranya akurasi, *recall*, dan *precision*, dan F-Measure. Dengan menggunakan *confusion matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	True Positif (TP)	False Negatif (FN)
Negatif	False Positif (FP)	True Negatif (TN)

a. Accuracy

$$accuracy = \frac{TP+TN}{All} \times 100\%$$

b. Precision

$$precision = \frac{TP}{(TP+FP)}$$

c. Recall

$$recall = \frac{TP}{(TP+FN)}$$

d. F Measure

F Measure

$$= \frac{2 \times recall \times precision}{(recall + precision)}$$

9. Word cloud

Word cloud (disebut juga *text cloud* atau *tag cloud*) adalah representasi visual baru dari data teks, biasanya digunakan untuk memvisualisasikan teks yang muncul dari sebuah web / metadata.

METODE PENELITIAN

1. Sumber Data

Data yang digunakan dalam penelitian ini adalah seluruh *tweet* menggunakan kata “Netflix blokir” di *Twitter* pada bulan 1 Januari 2020 sampai dengan 30 April 2020 sebanyak 1771 *tweet*.

2. Variabel Penelitian

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).

3. Analisis Data

Tahapan analisis yang digunakan dalam penelitian ini adalah :

1. *Scrapping* data *tweet* dengan keyword “Netflix blokir”
2. Melakukan tahap *text preprocessing*, yaitu : *cleaning*, *case folding*, *tokenizing*, *stopwords*, dan *stemming*
3. Melakukan pelabelan sentimen pada *tweet*
4. Membagi data *testing* dan data *training*
5. Klasifikasi data menggunakan metode *Naïve Bayes Classifier*
6. Klasifikasi data menggunakan metode

Support Vector Machine
kernel RBF

7. Melakukan pengujian performa metode NBC dan SVM berdasarkan tingkat *accuracy*, *precision*, *recall*, dan *F Measure*
8. Melakukan visualisasi *tweet* dengan *word cloud*
9. Interpretasi hasil analisis dan menarik kesimpulan

HASIL DAN PEMBAHASAN

1. Text Preprocessing dan Karakteristik Data

Text	<i>tweet</i>
Preprocessing	
<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*.

2. Pelabelan Data

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*.

Sentimen	Skor	Tanggapan / Tweet
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)

3. Term Frequency-Inverse Document Frequency (TF-IDF)

Data teks terlebih dahulu akan diubah ke dalam bentuk vector menggunakan teknik *Term Frequency-Inverse Document Frequency (TF-IDF)*. Untuk 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:

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

Q	Tf			IDF	W	
	d_1	d_2	d_f		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

4. Metode Klasifikasi

A. Naïve Bayes Classifier

Langkah pertama dalam mengklasifikasikan data tweet adalah melatih algoritma data *training*. Data *training* dari setiap *tweet* yang telah dilakukan praproses dilatih menggunakan software R. Selanjutnya melakukan data *testing* untuk mengetahui performa algoritma dari data *training*. Pembagian data *training* dengan data *testing* sebesar 70% dan 30%. Klasifikasi dengan metode *Naïve Bayes Classifier* menghasilkan

Support Vector	86.04%
Machine	

Berdasarkan nilai acc dan F -Measure terhadap $Support Vector$ Machine Kernel RBF memiliki lebih bagus dari pada $Classifier$ dalam menentukan sentimen.

6. Visualisasi Word

Aktual	Prediksi		Jumlah
	Positif	Negatif	
Positif	318	26	344
Negatif	48	138	186
Accurasy	86.04 %		530

Berdasarkan hasil perhitungan menggunakan *confusion matrix* pada algoritma *Support Vector Machine Kernel*

Naïve Bayes untuk Netflix blokir diperoleh nilai *accuracy* sebesar 86.04%.

5. Perbandingan Kinerja Metode NBC dan SVM Kernel RBF

Klasifikasi	Accuracy	Precision	Recall	F-Measure
Naïve Bayes Classifier	75.47%	79.11%	83.77%	81.37%
Support Vector Machine	86.04%	86.49%	92.44%	89.36%

Berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F-Measure* terlihat bahwa metode *Support Vector Machine* menggunakan Kernel RBF memiliki performa dan akurasi lebih bagus dari pada metode *Naïve Bayes Classifier* dalam mengklasifikasikan data sentimen.

6. Visualisasi Word Cloud



Word Cloud “Netflix Blokir” Sentimen Positif

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.



Word Cloud “Netflix Blokir” Sentimen Negatif

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.

KESIMPULAN

1. Hasil ketepatan klasifikasi menggunakan *Naive Bayes Classifier* dalam

mengklasifikasikan sentimen pengguna *Twitter* mengenai Netflix yang diblokir diperoleh akurasi sebesar 75.47%. Sedangkan pada metode *Support Vector Machine* menggunakan kernel RBF diperoleh akurasi sebesar 86.04%

2. Secara keseluruhan perbandingan performa metode *Naive Bayes Classifier* dan *Support Vector Machine* menunjukkan hasil performa *Support Vector Machine* lebih baik dalam mengklasifikasi sentimen. Hal ini dibuktikan oleh nilai akurasi *Support Vector Machine* lebih tinggi dibandingkan dengan nilai akurasi *Naive Bayes Classifier*.

3. 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.

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.
- Christianini, N., & John S, T. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press.
- Cnbc.com. 2020. Netflix: Diburu Sri Mulyani, Diblokir Telkom & 'Dibidik' MUI. <https://www.cnbcindonesia.com/tech/20200123142230-37-132199/netflix-diburu-sri-mulyani-diblokir-telkom-dibidik-mui>. Diakses pada tanggal 28 Juni 2020.
- Detikinet. 2020. Jumlah Pengguna Twitter meningkat, tapi ..., <https://inet.detik.com/cyberlife/d-5001786/jumlah-pengguna-Twitter-meningkat-tapi> . Diakses pada tanggal 28 Juni 2020
- Dragut, E., Fang, F., Sistla, P., Yu, S. & Meng, W. 2009. Stop Word and Related Problems in Web Interface Integration. *VLDB Endowment*.
- Falahah & Nur, D. D. A. (2015). Pengembangan Aplikasi Sentimen Analysis Menggunakan Metode Naïve Bayes. Seminar Nasional Sistem Informasi Indonesia, 335-340
- Feldman, R & Sanger, J. 2007. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press .
- 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.
- Hotho, A., Nurnberger, A., & Paass, G. (2005). A Brief Survey of Text Mining. Kassel: University of Kassel.
- Informatikologi. 2016. *Text Preprocessing*. [https://informatikologi.com/text-preprocessing/#:~:text=Salah%20sa%20implementasi%20dari%20text,%2C%20dan%20\(4\)%20stemming](https://informatikologi.com/text-preprocessing/#:~:text=Salah%20sa%20implementasi%20dari%20text,%2C%20dan%20(4)%20stemming). Diakses tanggal 30 Juni 2020
- Katadata. 2020. Rata-rata Waktu yang Dhabiskan Pengguna Internet Umur 16-24 Tahun. <https://databoks.katadata.co.id/data-publish/2020/02/26/indonesia-habiskan-hampir-8-jam-untuk-berinternet>. Diakses tanggal 30 Juni 2020
- Khotimah, Nur & Rochdi Wisono. 2019. *Sentimen Analysis of E-Commerce Brand Review Using Multinomial Text Naïve Bayes*. Skripsi: Program Studi Statistika Fakultas MIPA,

- Universitas Muhammadiyah Semarang.
- Kompas.com. 2016. Akhirnya Masuk Indonesia, Netflix Itu Apa?. <https://yogyakarta.kompas.com/read/2016/01/07/13085347/Akhirnya.Masuk.Indonesia.Netflix.Itu.Apa>. Diakses pada tanggal 28 Juni 2020.
- Lim, S. Y., Song, M.H., & Lee, S.J. 2006. Ontology-based automatic classification of web documents. Springer-Verlag, 690-700.
- Liu, Bing. 2012. *Sentimen Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- Nasukawa, T. & Yi, J., 2003. *Sentimen Analysis: Capturing Favorability Using Natural Language Processing*. In Proceedings of the 2nd International Conference on Knowledge Capture. pp. 70-77.
- Netflix. Di Wikipedia, *Ensiklopedia Bebas*. <https://id.wikipedia.org/wiki/Netflix>. Diakses tanggal 26 Juni 2020
- Nugroho, Kuncahyo Setyo. 2019. Dasar Text Preprocessing dengan Python. <https://medium.com/@ksnugroho/dasar-text-preprocessing-dengan-python-a4fa52608ffe>. Diakses tanggal 30 Juni 2020.
- Oktarialdi, B.C. 2014. Perbandingan Verifikasi Tanda Tangan Dengan Menggunakan Jaringan Saraf Tiruan Backpropagation dan Support Vector Machine. Tesis. Program Studi S2 Ilmu Komputer FMIPA UGM Yogyakarta.
- 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.
- Prasetyo,Eko. 2014. Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta : ANDI.
- Saraswati, N. S. (2011). Text Mining dengan Metode *Naive Bayes Classifier* dan *Support Vector Machines* untuk *Sentimen Analysis*. Tesis: Program Studi Teknik Elektro, Program Pasca Sarjana, Universitas Udayana.
- Setiabudi, Nur Andi. 2015. Membuat Word Cloud dengan R. <https://www.nurandi.id/blog/membuat-word-cloud-dengan-r/>. Diakses tanggal 30 Juni 2020
- Siang, J. J. (2005). Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB. Yogyakarta: ANDI.
- Suyanto. 2017. Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung:Informatika Bandung
- Tala, Fadillah Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Amsterdam: Institute for Logic, Language and Computation, Universiteit van Amsterdam The Netherlands.
- Troussas, C. et al., 2013. *Sentimen analysis of Facebook statuses using Naive Bayes classifier for language*

- learning. In IISA 2013. IEEE, pp. 1–6.
- Twitter. (2020, Juni 28). *Company*. Retrieved from https://about.Twitter.com/en_us/company.html
- Twitter. Di Wikipedia, *Ensiklopedia Bebas*.
<https://id.wikipedia.org/wiki/Twitter>
. Diakses tanggal 26 Juni 2020
- 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

