

# Sentimen Analisis *Twitter* Terhadap Isolasi Diri Masyarakat Indonesia Akibat Dampak *Covid-19*

Resi Taufan, Tri Rivanie, Sri Rahayu, Windu Gata

**Abstract**— The rise of the Corona or Covid-19 virus in Indonesia has resulted in the government recommending that Indonesian people carry out self-isolation and social distancing. This has caused many public responses, especially in the social media Twitter about these recommendations. In this research an analysis of community sentiment will be carried out on the impact felt by the community regarding self-isolation. There are several steps to do sentiment analysis, including the stages of data collection, data preprocessing, data validation and testing done with the rapidminer tool by using the TF-IDF feature to weight each word. In this study each tweet is classified into positive and negative sentiment. The Naive Bayes method in sentiment classification has a pretty good accuracy to classify sentiments.

**Index Terms**— Covid-19; Self Isolation; Sentiment Analyst; TF-IDF;

**Abstrak**— Maraknya penyebaran virus Corona atau Covid-19 di Indonesia mengakibatkan pemerintah menganjurkan agar masyarakat Indonesia melakukan isolasi diri dan *social distancing*. Hal ini menyebabkan banyaknya tanggapan publik khususnya di media sosial *twitter* tentang anjuran tersebut. Dalam penelitian ini akan dilakukan analisis sentimen masyarakat terhadap dampak yang dirasakan masyarakat mengenai isolasi diri. Ada beberapa tahap untuk melakukan analisis sentimen, diantaranya adalah tahap pengumpulan data, *preprocessing* data, validasi data serta pengujian dilakukan dengan tools *rapidminer* dengan menggunakan fitur TF-IDF untuk melakukan pembobotan pada setiap kata. Pada penelitian ini setiap *tweet* diklasifikasikan menjadi sentiment positif dan negative. Metode Naive Bayes dalam klasifikasi sentimen memiliki akurasi yang cukup baik untuk melakukan klasifikasi sentimen.

**Kata Kunci**— Covid-19; Isolasi Diri; Sentimen Analisis; TF-IDF;

Resi Taufan, mahasiswa program studi Ilmu Komputer, STMIK Nusa Mandiri, email : [resitauf1508@nusamandiri.ac.id](mailto:resitauf1508@nusamandiri.ac.id).

Tri Rivanie, mahasiswa program studi Ilmu Komputer, STMIK Nusa Mandiri, email : [14002396@nusamandiri.ac.id](mailto:14002396@nusamandiri.ac.id).

Sri Rahayu, dosen program studi Sistem Informasi, STMIK Nusa Mandiri, email : [srirahayu.ry@nusamandiri.ac.id](mailto:srirahayu.ry@nusamandiri.ac.id).

Windu Gata, dosen program studi Ilmu Komputer, STMIK Nusa Mandiri, email : [windu@nusamandiri.ac.id](mailto:windu@nusamandiri.ac.id).

## I. PENDAHULUAN

Belakangan ini masyarakat dunia diresahkan oleh Virus Corona atau *Covid-19* yang penyebarannya termasuk paling besar dibandingkan dengan SARS (*Severe Acute Respiratory Syndrome*) dan *Middle East respiratory syndrome coronavirus (MERS-CoV)*, sampai dengan Kamis 12 Maret 2020 tercatat sudah menginfeksi 126.061 orang di 118 negara di dunia. Berdasarkan perhitungan situs pelaporan daring *Worldometers*, sekitar 67.064 orang dinyatakan sembuh dan 4.616 meninggal dunia (CNN Indonesia). Pemerintah Indonesia menganjurkan agar masyarakat untuk melakukan isolasi diri dan *social distancing*. Banyak masyarakat yang menanggapi kebijakan tersebut di sosial media dengan berbagai sentimen baik positif maupun negatif. Analisis sentimen atau *opinion mining* merupakan salah satu solusi mengatasi masalah untuk mengelompokkan opini atau review menjadi opini positif atau negatif secara otomatis [1].

Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas opini dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari *text mining*, dalam hal ini adalah klasifikasi dokumen [2].

Penggunaan media sosial seperti *Facebook*, *Twitter*, *Pinterest*, dan *Youtube* pada era abad informasi ini mengalami kenaikan yang sangat drastis dibandingkan dengan tahun – tahun sebelumnya. Data dari tahun 2015 saja memaparkan pengguna *Twitter* ada lebih dari 285 juta [3]. Meskipun begitu sebuah analisis teks tidak harus dengan media sosial tertentu, pemilihan pada media sosial *Twitter*, karena kelebihan *Twitter* memakai karakter dibawah 140, format pesan yang informal, teks pesan yang sangat banyak setiap harinya, berasal dari

individu dengan latar belakang yang sangat bervariasi, dan mudah digunakan [4].

## II. LANDASAN TEORI

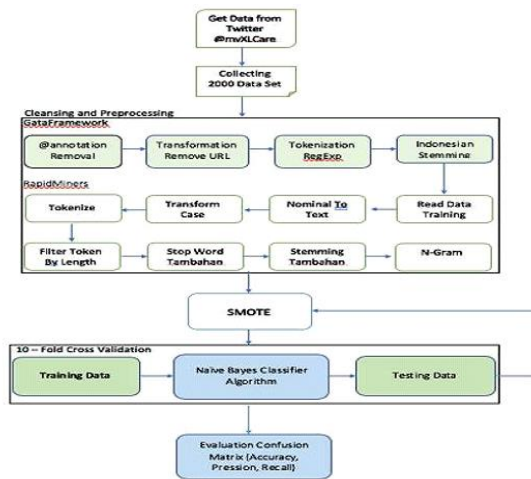
Ada beragam teknik klasifikasi dokumen, di antaranya adalah *Naive Bayes Classifier*, *Decision Trees*, dan *Support Vector Machines*. Salah satu metode yang paling populer digunakan dalam pengklasifikasian dokumen sekarang ini adalah metode *Naive Bayes Classifier* [5]. Hal serupa juga diungkapkan oleh [6] dalam penelitiannya, yaitu metode *Naive Bayes Classifier* memiliki beberapa kelebihan antara lain, sederhana, cepat dan berakurasi tinggi. Penelitian mengenai analisis sentimen telah banyak dilakukan sebelumnya. Penelitian tentang analisis sentimen yang menggunakan dataset dari jejaring sosial *Twitter* dilakukan oleh [7]. Mereka melakukan analisis sentimen terhadap media jejaring sosial *Twitter* dengan menggunakan beberapa teknik klasifikasi. Penelitian sentimen analisis dengan metode *Naive bayes* memiliki akurasi diatas 75 % algoritma ini sering digunakan dalam analisis sentimen teks [8].

Selain metode *Naive bayes* pada analisis text mining dilakukan pembobotan kata pada masing-masing *tweet* dengan memanfaatkan *Term Frequency - Inverse Document Frequency (TF-IDF)*, Penerapkan *TF-IDF* untuk text mining sangat membantu pengguna. untuk mendapatkan informasi pada kumpulan dokumen. Dengan format file txt berdasarkan kata kunci yang dimasukan oleh pengguna pada system [9].

Penelitian selanjutnya yang menjadi acuan penulis dalam menyusun penelitian ini adalah penelitian yang dilakukan oleh [10]. Pada penelitian tersebut dilakukan analisis sentimen menggunakan metode *naive bayes* dalam penentuan polaritas sentimen. Hasil dari penelitian tersebut menunjukkan akurasi yang cukup tinggi untuk metode *Naive Bayes Classifier*. Atas dasar hal ini, penulis bermaksud menerapkan metode *Naive Naves Classifier* untuk melihat sentimen masyarakat di media *Twitter* terhadap isolasi diri akibat dampak dari *covid-19*.

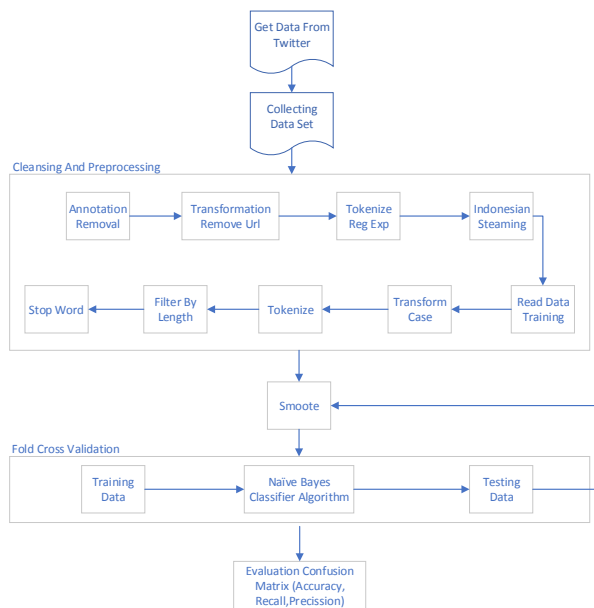
## III. METODE PENELITIAN

Rancangan penelitian dilakukan dengan proses data *retrieval* dari *twitter* sebanyak 525 *tweet* dengan *query* isolasi diri, kemudian dilakukan *crowling* dengan tools *rapidminer*.



Gambar 1. Metodologi Penelitian Klasifikasi Sentimen Analisis[10]

Pada penelitian ini rancangan yg dibangun hampir sama dengan penelitian sebelumnya [10] hanya saja proses *cleansing* dan *preprocessing* dilakukan dengan *rapidminer* tidak dengan *framework* tambahan.



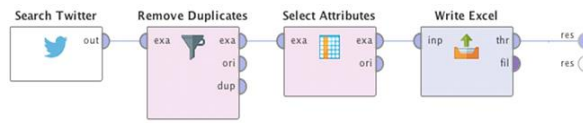
Gambar 2. Metodologi Penelitian Klasifikasi Sentimen Analisis.

## IV. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dari pengolahan data awal , preprocessing dan validasi. Berikut adalah tahapannya :  
a. Pengolahan Data Awal

Data yang digunakan pada penelitian bersumber dari *twitter* dengan cara mengkoneksikan *raidminer* dengan *API twitter* sebanyak 2000 *tweet*. *Crawling* data menggunakan aplikasi *rapidminer* dengan *query* ‘Isolasi Diri’. *Twitter* memungkinkan redudansi data dimana ada beberapa komentar yang di *retweet* sehingga perlu adanya *Remove Duplicate*, dari 2000 data terpilih 525 data unik atau tidak duplikat, kemudian dipilih data yang akan digunakan sebagai

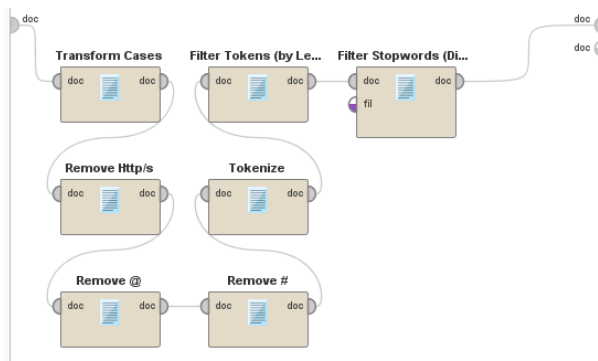
text, data ini berisi komentar positif dan negatif, setelah itu dilakukan palebelan positif atau negatif pada setiap komentar.



Gambar 3. Proses Data Retrieval dari Twitter[10]

b. Preprocessing

Tahapan ini merupakan tahap awal untuk mengolah data teks menjadi analisa sentimen. Proses dilanjutkan dengan menghitung probabilitas masing-masing kata dalam tweet berdasarkan data training menggunakan fitur TF-IDF. Berikut tahapan preprocessing yang digunakan :



Gambar 4. Preprocessing pada tool rapidminer

1) Transform Cases

Pada proses ini, rapidminer akan mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf. Setelah melalui tahap ini, seluruh isi komentar akan menjadi non kapital.

Tabel 1. Hasil Transform cases

Text	Transform Cases
Berbelanja keperluan primer untuk keperluan isolasi diri hingga 2021 <a href="https://t.co/UHqrR8z0yC">https://t.co/UHqrR8z0yC</a>	berbelanja keperluan primer untuk keperluan isolasi diri hingga 2021 <a href="https://t.co/uhqrr8z0yc">https://t.co/uhqrr8z0yc</a>

2) Tokenization

Proses tokenization berfungsi untuk menghilangkan tanda baca, simbol dan karakter yang bukan berupa huruf pada setiap komentar. Proses tokenization pada penelitian ini seperti terlihat pada gambar 4 yaitu menghilangkan kata yang diawali http/https (Remove Http/s), menghilangkan karakter @ (Remove @) dan menghilangkan karakter # (Remove #).

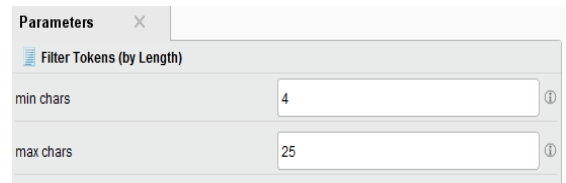
Tabel 2. Hasil Tokenize

Text	Tokenize
@null Tren Selebgram Pamer Isolasi Diri di Instagram, di Bali Hingga	@null Tren Selebgram Pamer Isolasi Diri di Instagram, di Bali Hingga Vila Mewah

Vila Mewah Banyak <a href="https://t.co/6Sx1lgTpOC">https://t.co/6Sx1lgTpOC</a>	Banyak
@ganjarpranowo Ngapnten pak, didesa saya paklurahnya menganjurkan untuk isolasi diri dirumah selama 14 hari nanti setiap harinya dipantau dari puskesmas. Ngoten pak	ganjarpranowo ngapnten pak, didesa saya paklurahnya menganjurkan untuk isolasi diri dirumah selama 14 hari nanti setiap harinya dipantau dari puskesmas ngoten pak
Sejatinya pemerintah setempat sudah menyediakan gedung untuk mengisolasi diri, namun dinilai tak layak. #publisherstory <a href="https://t.co/iFsi2Jt3a2">https://t.co/iFsi2Jt3a2</a>	sejatinya pemerintah setempat sudah menyediakan gedung untuk mengisolasi diri, namun dinilai tak layak. publisherstory

3) Filter Token (by Length)

Pada tahap ini dilakukan pemilihan token dengan ukuran panjang minimal 4 karakter (huruf), walaupun beberapa diantaranya adalah stopwords. Jika kata yang kurang dari 4 huruf namun termasuk dalam stopwords akan tetap dibuang, karena tahap selanjutnya adalah pembuangan stopwords. Proses filter token dapat dilihat pada Gambar 5.



Gambar 5. Filter by length pada tool rapidminer

Tabel 3. Hasil filter token (by length)

text	Filter Token (By Length)
kgm bgt bro gmn kl kita ketemu nnt malem trs isolasi diri brg	ketemu malem isolasi diri

4) Stopwords

Pada tahap ini akan menyempurnakan tahap filter token by length. Kata-kata yang dibuang yaitu kata-kata yang tidak relevan seperti “untuk”, “adalah”, ”sebuah”, “oleh”, “atau”, “pada”, “dalam”, “sebagai”, “menjadi”, “itu” dan sebagainya.

Tabel 4. Hasil Filter Token (By Length)

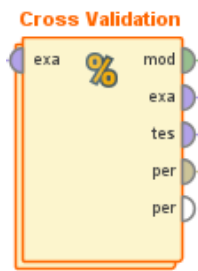
text	Filter Token (By Length)
Koruptor berada di Lapas Sukamiskin orang yang telah melakukan isolasi diri yang paling sempurna ditengah wabah Covid -19, sehat2 semua , paling hanya satu mereka tidak bisa piknik.	Koruptor berada Lapas Sukamiskin melakukan isolasi diri paling sempurna ditengah wabah Covid -19, sehat2 semua , paling hanya satu mereka tidak bisa piknik.

5) *Stemming*

Pada tahap *stemming*, kata dengan kata hubung diubah menjadi kata dasar untuk tweet dalam bahasa Indonesia.

c. Validasi

Menguji *dataset* dengan komposisi data pengujian dan *parameter* pelatihan yang berbeda dilakukan dengan mendistribusikan data pelatihan dan menguji data dengan komposisi berdasarkan distribusi data dalam validasi silang *k-fold* dengan nilai  $k = 2$  hingga  $k = 10$ . Tes ini bertujuan untuk menentukan nilai *k-fold* optimal dalam proses klasifikasi opini dalam kasus penelitian ini.



Gambar 6. Operator *Cross validation K-fold*

d. Evaluasi

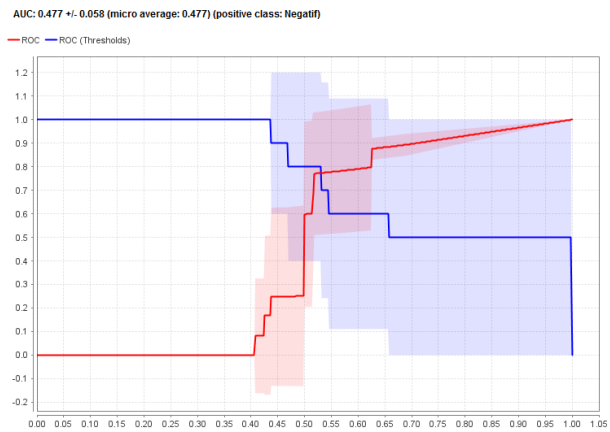
Hasil dari implementasi model *Naive Bayes* tersebut akan didapat seluruh *rule*. *Rule* yang dihasilkan akan digunakan sebagai dasar prediksi nilai yang akan dilakukan. Sebelumnya *rule* tersebut harus dievaluasi dan divalidasi sehingga diketahui seberapa akurat hasil prediksi yang akan dilakukan. Pada tahap ini proses evaluasi menggunakan metode *confusion matrix*. Model *confusion matrix* akan membentuk matriks yang terdiri dari *accuracy*, *true positif* dan *true negatif*, *recall* dan *precision*. Berikut merupakan hasil dari *confusion matrix* pada algoritma *naive bayes classifier*

accuracy: 67.44% +/- 4.23% (micro average: 67.45%)

	true Positif	true Negatif	class precision
pred. Positif	163	51	76.17%
pred. Negatif	158	270	63.08%
class recall	50.78%	84.11%	

Gambar 7. *Accuracy, True Positif dan True Negatif, Recall dan Precision.*

Hasil akurasi pengujian *Naive Bayes* untuk analisis sentimen pada media sosial *twitter* memperoleh hasil sebesar 67,44 % sedangkan untuk nilai *AUC* nya sebesar 0.477 yang dapat dilihat pada Gambar 8.



Gambar 8. *AUC Process NBC*

V. KESIMPULAN

Hasil analisa dalam penelitian ini penulis menyimpulkan bahwa sentimen masyarakat indonesia terhadap isolasi diri sebagai dampak dari *Covid-19* dengan menggunakan metode *Naive Bayes* dan fitur *TF-IDF* pada media sosial *twitter* menghasilkan akurasi yang cukup baik yaitu 67,44 %

REFERENSI

- [1] E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," vol. 4, 2016.
- [2] F. Nurhuda and S. W. Sihwi, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," vol. 2, no. 2, 2014.
- [3] S. Alim, "Analysis of tweets related to cyberbullying: Exploring information diffusion and advice available for cyberbullying victims," *Int. J. Cyber Behav. Psychol. Learn.*, vol. 5, no. 4, pp. 31–52, 2015, doi: 10.4018/IJCBPL.2015100103.
- [4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," 2008.
- [5] S. Natalius, "Makalah II2092 Probabilitas dan Statistik-Sem. I Tahun," no. 3, 2010.
- [6] R. Mccue, "A Comparison of the Accuracy of Support Vector Machine and Naive Bayes Algorithms In Spam Classification," pp. 1–17, 2009.
- [7] R. Parikh and M. Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques," pp. 1–18, 2009.
- [8] A. Fauzi, A. N. Rais, M. F. Akbar, and W. Gata, "ANALISIS SENTIMEN BERINTERNET PADA MEDIA SOSIAL NAIVE BAYES," pp. 46–54, 2018.
- [9] M. Nurjannah and I. Fitri Astuti, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman," *J.*

*Inform. Mulawarman*, vol. 8, no. 3, pp. 110–113, 2013.

- [10] D. D. Saputra, W. Gata, N. K. Wardhani, and K. Sakho, "Optimization Sentiments of Analysis from Tweets in myXLCare using Naïve Bayes Algorithm and Synthetic Minority Over Sampling Technique Method Optimization Sentiments of Analysis from Tweets in myXLCare using Naïve Bayes Algorithm and Synthetic Minority Ov," 2020, doi: 10.1088/1742-6596/1471/1/012014.