

## ANALISIS SENTIMEN TERHADAP KINERJA MENTERI KESEHATAN INDONESIA SELAMA PANDEMI COVID-19

Tri Rivanie<sup>1</sup>, Rangga Pebrianto<sup>2</sup>, Taopik Hidayat<sup>3</sup>, Achmad Bayhaqy<sup>4</sup>, Windu Gata<sup>5</sup>, Hafifah Bella Novitasari<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Ilmu Komputer, STMIK Nusa Mandiri Jakarta

Jl. Kramat Raya No.18, Jakarta Pusat, Jakarta 10420

Email: [14002326@nusamandiri.ac.id](mailto:14002326@nusamandiri.ac.id)<sup>1</sup>, [14002396@nusamandiri.ac.id](mailto:14002396@nusamandiri.ac.id)<sup>2</sup>,

[Taopik.Tao@nusamandiri.ac.id](mailto:Taopik.Tao@nusamandiri.ac.id)<sup>3</sup>, [achmad.acq@nusamandiri.ac.id](mailto:achmad.acq@nusamandiri.ac.id)<sup>4</sup>, [Windu@nusamandiri.ac.id](mailto:Windu@nusamandiri.ac.id)<sup>5</sup>, [hafifahbela@stmik-banisaleh.ac.id](mailto:hafifahbela@stmik-banisaleh.ac.id)<sup>6</sup>

### ABSTRACT

*The pandemic that occurred in Indonesia has not yet subsided and far from under control. Indonesian Ministry of Health is most appropriate person to responsible for providing an explanation of actual situation and extent to which state has handled it. However, he has rarely appeared in public lately to explain about handling of Covid-19 pandemic. In response, many people are pros and cons come to give their opinions and feedback. The increasing use of internet during pandemic, especially on social media, where one of them is Twitter, which is a means of expressing opinions. Posting tweets is a community habit to assess or respond to events, as well as represent public's response to an event, especially Ministry of Health steps and policies in handling and breaking chain of Covid-19 pandemic.*

*The tweet posts were taken only in Indonesian-language and also related to performance of Government, especially Ministry of Health. After that, a label is given so that sentiment of tweets is known. To test results of these sentiments, an algorithm is used by comparing two methods of Support Vector Machine (SVM) and Naïve Bayes (NB). Validation was carried out using k-Fold Cross Validation to obtain an accuracy value. The results show that accuracy value for NB algorithm is 66.45% and SVM algorithm has a greater accuracy value of 72.57%. So it can be seen that SVM algorithm managed to get the best accuracy value in classifying positive comments and negative comments related to sentiment analysis towards Ministry of Health.*

**Keywords**—Support Vector Machine, Naïve Bayes, Analisis sentimen, K-Fold Cross Validation

### ABSTRAK

Pandemi yang terjadi di Indonesia belum juga mereda dan terkendali. Menteri kesehatan Indonesia merupakan sosok yang dianggap paling tepat untuk bertanggung jawab memberi penjelasan mengenai situasi sebenarnya dan sejauh apa penanganan yang sudah dilakukan oleh negara. Namun, beliau sangat jarang tampil akhir-akhir ini di hadapan publik untuk menjelaskan mengenai penanganan pandemi Covid-19. Menanggapi hal tersebut, banyak masyarakat yang pro maupun kontra ikut memberikan pendapat serta tanggapan. Meningkatnya penggunaan internet selama masa pandemi terutama pada media sosial, dimana salah satu wadahnya yaitu *twitter* yang menjadi sarana menuangkan pendapat. Memposting *tweet* merupakan kebiasaan masyarakat untuk menilai atau menanggapi peristiwa, serta merepresentasikan tanggapan masyarakat terhadap suatu peristiwa terutama langkah dan kebijakan Kementerian Kesehatan dalam penanganan dan pemutusan rantai pandemi Covid-19.

Postingan *tweet* tersebut diambil yang berbahasa Indonesia dan juga berkaitan dengan kinerja Pemerintah, khususnya Kementerian Kesehatan. Setelah itu, dilakukan pemberian label agar diketahui sentimen dari *tweet* tersebut. Untuk menguji hasil sentimen tersebut, dilakukan menggunakan algoritma dengan membandingkan dua metode *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB). Dilakukan validasi menggunakan *k-Fold Cross Validation* untuk memperoleh nilai akurasi (*accuracy*). Dari hasilnya menunjukkan bahwa nilai akurasi untuk algoritma NB sebesar 66,45% dan algoritma SVM memiliki nilai akurasi lebih besar yaitu 72,57%. Maka dapat diketahui bahwa algoritma SVM berhasil mendapatkan nilai akurasi yang terbaik dalam mengklasifikasikan komentar positif dan komentar negatif terkait dengan analisis sentimen terhadap Kementerian Kesehatan.

**Kata Kunci**—*Support Vector Machine, Naïve Bayes, Analisis sentimen, K-Fold Cross Validation*

## I. PENDAHULUAN

Saat ini seluruh dunia sedang direpotkan dengan virus baru *Covid-19*, *South China Morning Post* melaporkan kasus *Covid-19* pertama di dunia ditemukan pada 17 November 2019 di Cina. Infeksi ini ditemukan pada pria berusia 55 tahun dari Provinsi *Hubei*. Sementara kasus *Covid-19* pertama di Indonesia diumumkan pada 2 Maret 2020 atau sekitar 4 bulan setelah ditemukan kasus *Covid-19* pertama di Cina. Total kasus corona di Indonesia pada 2 Maret sebanyak dua kasus. Kondisi pandemi di Indonesia saat ini belum juga terkendali, data dan angka jelas menunjukkan itu. Berdasarkan data dari pemerintah di *Covid19.go.id*, kurva kasus positif *Covid-19* di Indonesia terus bergerak naik dan belum ada tanda-tanda penurunan. DKI Jakarta menjadi wilayah dengan kasus positif *Covid-19* terbanyak yakni 50.671,

diikuti Jawa Timur 37.093, Jawa Tengah 16.508 dan Jawa Barat 13.668 kasus.

Dalam hal ini sosok Menteri Kesehatan Dr. dr. Terawan Agus Purhanto, Sp.Rad kembali menjadi sorotan. Menteri Kesehatan dianggap menjadi sosok yang bertanggung jawab dalam memberi penjelasan mengenai situasi sebenarnya dan sejauh apa penanganan yang sudah dilakukan oleh Negara. Karena beliau yang memiliki wewenang akses anggaran dan pemberi arahan. Namun kini Menteri Kesehatan Terawan Agus Purhanto sangat jarang tampil di hadapan publik untuk menjelaskan mengenai penanganan pandemi *Covid-19*. Jarang terlihatnya Menkes Terawan di media mulai jadi sorotan sejak jurnalis sekaligus presenter Najwa Shihab dalam acara "Mata Najwa" edisi "Menanti Terawan" bermonolog dengan kursi kosong seolah-olah ada Terawan yang tengah duduk di sana. Dalam monolognya, Najwa mengungkap kegelisahan masyarakat yang

sudah jarang melihat sosok Menkes Terawan di media sejak pandemi *Covid-19* melanda Indonesia. *Rapid test* yang belum mencapai target dan perlindungan terhadap tenaga kesehatan sebagai garda terdepan belum maksimal, angka kematian tenaga kesehatan sangat tinggi dan masih terus naik, dimana Kementerian kesehatan seharusnya menjadi pelindung dan pembela tenaga kesehatan. Selain itu masih saja ada disparitas antara data pusat dan daerah, padahal data tersebut sangat krusial dalam menentukan kebijakan. Bahkan data menunjukkan bahwa Gedung kementerian menjadi *cluster* perkantoran terbesar di Jakarta. membuat opini masyarakat terhadap Menteri kesehatan bahwa kini beliau telah menyadari bahwa Indonesia telah kecolongan dalam penanganan langkah diawal yang seharusnya bisa lebih tanggap. Menurut Najwa, kehadiran Terawan sangat diperlukan untuk menjelaskan kepada publik terkait situasi pandemi *Covid-19* di Indonesia.

Menanggapi hal tersebut banyak masyarakat yang pro dan kontra ikut memberikan pendapat dan tanggapan menyoroti Kinerja Pemerintah terkait *Covid-19*, terutama kinerja Kementerian kesehatan, kemudian memposting di media sosial. Media sosial telah menjadi bagian dari masyarakat global, di mana

orang-orang dari seluruh dunia membagikan pendapat dan pandangan mereka secara bebas di blog pribadi, *Facebook*, *YouTube*, *Twitter* dan lain-lain. salah satu media sosial yang berkembang pesat penggunaannya di Indonesia adalah *Twitter*. *Twitter* adalah *platform* media sosial dianggap paling populer untuk berbagi masalah politik dan masalah social [1]. Saat ini *Twitter* merupakan sebuah indikator yang baik untuk memberikan pengaruh dalam penelitian [2]. Banyaknya *tweet* dari pengguna *twitter* hari ke hari bertambah banyak, data yang berkembang di *Twitter* dianggap lebih diminati karena dirasa cukup mudah dan simpel dalam merepresentasikan apa yang berkembang di masyarakat pada umumnya.

Informasi yang tersimpan cenderung berupa teks dilakukan proses ekstraksi informasi yang berpola (*text mining*). *Text mining* adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. *Text mining* memiliki tujuan dan menggunakan proses yang sama dengan *Data Mining*, namun memiliki masukan yang berbeda. Masukan untuk *text mining* adalah data yang tidak (atau kurang) terstruktur, seperti dokumen *word*, *PDF*, kutipan teks dan lain-lain. Sedangkan masukan untuk *Data Mining* adalah data yang terstruktur [3].

Analisis sentimen yang merupakan bagian dari *opinion mining* [4]. Analisis sentimen adalah proses perbandingan untuk mengungkap dan mengategorikan pendapat yang diungkapkan dalam teks, untuk menentukan apakah tanggapan terhadap topik tertentu positif, negatif, atau netral [5]. Analisis sentimen dapat sebagai solusi dari masalah yang sudah dipaparkan sebelumnya [6]. Sentimen terutama mengacu pada perasaan, emosi, pendapat atau sikap. Dengan meningkatnya *World Wide Web*, orang sering mengungkapkan perasaan mereka melalui internet melalui media sosial, *blog*, *rating* dan *reviews*. Karena ini meningkatkan data tekstual, ada kebutuhan untuk menganalisa konsep mengekspresikan sentimen dan menghitung wawasan untuk menjelajahi bisnis [7].

Ada beberapa pendekatan yang berbeda yang dapat digunakan untuk permasalahan analisis sentimen, seperti *Naive Bayes* [9][10], dan *Bayesian Networks* [10][11], *K-Nearest Neighbor* [12][13], *Maximum Entropy* [14][15], *Neural Network* [16], [17], *Support Vector Machines* [6][18][19][20] dan lain-lain.

Pada penelitian ini algoritma yang akan digunakan adalah *Support Vector Machine* (SVM) dan *Naive Bayes* (NB). Kedua metode ini sering dijadikan metode

dalam penelitian yang membahas sentiment analisis. Pemilihan metode klasifikasi SVM karena memiliki kemampuan generalisasi dalam mengklasifikasikan suatu *pattern*, tidak termasuk data yang dipakai dalam fase pembelajaran metode tersebut [21]. Algoritma SVM digunakan pada analisis sentimen dikarenakan algoritma ini bertujuan sebagai metode klasifikasi kedalam kategori positif dan negatif [6].

Sedangkan Pengklasifikasi *Naive Bayes* (NB) sangat sederhana dan efisien [22]. Selain itu, pengklasifikasi *Naive Bayes* adalah teknik *machine learning* yang populer untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain [23]. Pada penelitian sebelumnya metode *Naive Bayes* juga digunakan dalam Analisis Sentimen berinternet pada media sosial AMIK BSI Tegal di mana metode *Naive Bayes* digunakan untuk menentukan sentimen publik terhadap AMIK BSI Tegal yang disampaikan di *Instagram*, sehingga membantu usaha untuk melakukan riset *marketing* atas *public opinion* [9].

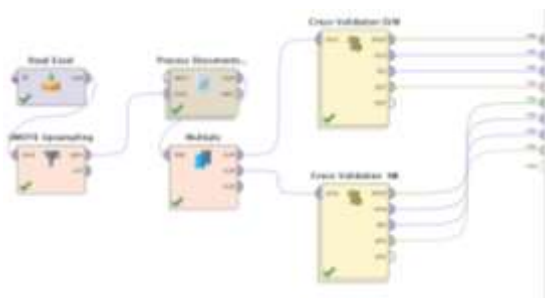
Oleh karena itu dalam penelitian ini digunakan metode SVM dan NB yang diharapkan sesuai dengan permasalahan. Dengan dilakukannya analisis sentimen ini, diharapkan dapat diketahui tanggapan masyarakat terhadap kepuasan kinerja

Kementerian Kesehatan, sehingga dapat dijadikan sarana dan masukan untuk membuat pertimbangan kebijakan Kementerian Kesehatan selanjutnya dalam penanggulangan penyebaran *Covid-19* di Indonesia.

## II. METODE PENELITIAN

Penelitian ini dimulai dengan melakukan analisa dan menentukan *query*, Penentuan *query* pada *Twitter* akan membantu dalam pengumpulan data *tweet* sehingga akan lebih terarah. *Query* yang dipilih dalam penelitian ini yaitu Menkes Terawan.

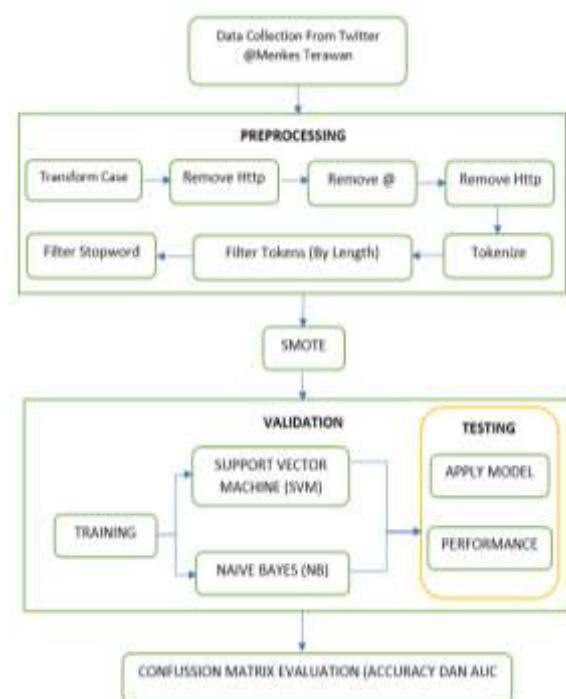
Data *Tweet* diambil dengan menggunakan metode *Crawling* dari sosial media *Twitter* menggunakan *Twitter API* menggunakan *Tools Rapidminer Studio* versi 9.6. Data yang diambil hanya *tweet* berbahasa Indonesia kemudian disimpan ke *Microsoft Excel* dimana setiap *query* terdiri dari 2.000 data. Model analisis data yang dibuat sebagaimana disajikan pada gambar 1.



**Gambar 1. Desain Model Komparasi Support Vector Machine dan Naive Bayes**

### 2.1. Model Penelitian

Langkah-langkah yang akan dilakukan pada penelitian ini dituangkan dalam sebuah model kerangka kerja penelitian. Tahapan-tahapan pada model kerangka kerja ini akan dijadikan acuan selama proses penelitian. Model kerangka kerja pada penelitian ini disajikan pada Gambar 2.



**Gambar 2. Kerangka Kerja Analisis Sentimen**

### 2.2. Teknik Pengumpulan data

Pada penelitian ini menggunakan data sekunder. Data yang digunakan untuk penelitian dapat berupa data primer maupun data sekunder. Data primer merupakan data yang diperoleh langsung dari sumber datanya, sedangkan data

sekunder merupakan data yang didapatkan dari sumber yang sudah ada.

### 2.3. Preprocessing

Pada tahapan ini dilakukan proses *preprocessing* yang diantaranya *transform case*, *remove http*, *remove annotation*, *remove hashtag*, *tokenize*, *stopword removal*.

#### a. Transform Case

*Transform case* bertujuan untuk mengubah semua teks menjadi huruf kecil. Hal ini bertujuan untuk menghindari masalah ketika dilakukan proses *tokenize*.

#### b. Remove Http

*Remove http* bertujuan untuk menghilangkan *URL* yang biasanya terdapat pada teks *tweet*.

#### c. Remove Annotation

*Remove Annotation* bertujuan untuk menghapus tanda *annotation* yang seringkali terdapat pada teks *tweet*. *Annotation* merupakan salah satu *noise* yang tidak memiliki arti.

#### d. Remove Hashtag

*Remove Hashtag* bertujuan untuk menghapus kata kunci *hashtag*.

#### e. Tokenize

*Tokenize* adalah proses yang bertujuan untuk memisahkan setiap kata di dalam teks menjadi sebuah entitas tunggal. Pemisahan kata tersebut dilakukan dengan mengidentifikasi karakter spasi atau tanda baca sebagai pemisah antar kata.

#### f. Stopword Removal

*Stopword Removal* adalah proses bertujuan menghilangkan kata-kata umum yang dianggap tidak memiliki makna karena kata-kata tersebut dianggap tidak dapat mewakili konten dokumen. Kata-kata umum tersebut diidentifikasi berdasarkan kata-kata yang terdapat dalam sebuah *dictionary*, dalam penelitian ini digunakan file "*stop-word-list-tala.txt*" yang berisi *stopwords* bahasa Indonesia.

### 2.4. SMOTE

Dalam pengumpulan Data terkait analisis sentimen setelah dilakukan *preprocessing* sangat mungkin terjadi *imbalance* data. *Imbalance* data merupakan kondisi dimana di antara kedua kelas memiliki selisih yang sangat signifikan. Pemodelan dengan algoritma yang tidak memperhatikan ketidakseimbangan data akan didominasi oleh kelas mayor dan tidak memperhatikan kelas minornya. Hal tersebut akan sangat mempengaruhi hasil model yang diperoleh [24].

Untuk mengatasi masalah ini, maka diperlukan *SMOTE* (*Synthetic Minority Oversampling Technique*), dengan melakukan *up-sampling* atau menambah jumlah data kelas minor agar setara dengan kelas mayor dengan membangkitkan data buatan.



### 2.5. Validation Data

Tahapan validation data dilakukan dalam proses *K-Fold Cross Validation* dimana evaluasinya sebanyak  $K=10$ . Algoritma yang digunakan untuk melakukan proses evaluasinya yaitu menggunakan *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB).

#### a. Support Vector Machine

*Support Vector Machine* (SVM) adalah metode klasifikasi yang bekerja dengan cara mencari *hyperplane*. SVM merupakan metode yang digunakan untuk klasifikasi dua kelas (*binary classification*). Pada perkembangannya, beberapa metode diusulkan agar SVM bisa digunakan untuk klasifikasi *multi-class* dengan cara mengombinasikan beberapa *binary classifier*.

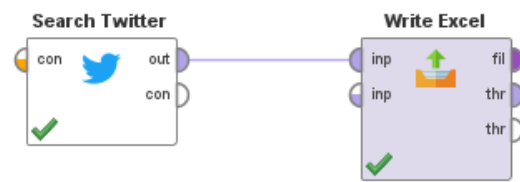
#### b. Naïve Bayes

*Naïve Bayes* (NB) adalah algoritma pembelajaran yang didasarkan pada *teorema bayes* dengan menggunakan asumsi yang kuat. *Teorema Bayes* merupakan teori tentang menemukan probabilitas tertinggi sesuatu berdasarkan data yang sudah ada.

## III. HASIL DAN PEMBAHASAN

### 3.1. Pengambilan Dataset

Proses Pengambilan data (*Crawling data*) menggunakan *Tools Rapidminer* diambil dari media sosial *Twitter*.



**Gambar 3. Proses *crawling twitter***

Data yang diambil adalah data terbaru dan terpopuler dan hanya *tweet* berbahasa Indonesia dengan jumlah *record* yang di *setting* sebanyak 2.000. Dari 2.000 data tersebut didapatkan 982 data *tweet*, kemudian dilakukan data *cleansing* untuk membersihkan data dari kata yang tidak diperlukan seperti *retweet*, *hashtag* dan untuk menghilangkan 672 duplikasi data yang menghasilkan 310 data *tweet*. Sehingga dataset yang siap diproses untuk *pre-processing* adalah sebanyak 310 baris.

### 3.2. Pelabelan kelas sentimen

Dari 310 data *tweet* kemudian ditentukan *class attribute*, *class attribute* yang akan digunakan dalam penelitian ini hanya menggunakan 2, yaitu positif dan negatif, karena pada Penelitian ini bertujuan menyimpulkan sistem analisis sentimen ini dapat mengenali kelemahan (berdasarkan komentar negatif) dan kelebihan (berdasarkan komentar positif) secara otomatis dengan baik.

### 3.3. Preprocessing Data

Data yang telah diperoleh baru akan bisa digunakan setelah melalui beberapa tahap sebelum proses pengujian. tahap ini

adalah tahap *Preprocessing* Data. Data akan diproses melalui beberapa tahap Berikut :



**Gambar 4. Proses Preprocessing Data**

Dalam Proses *document form Data* terdapat operator *Transform Case*, *remove HTTP*, *remove Annotation(@)*, *remove Hastag(#)*, *Tokenize*, *Filter Token By Length*, *Filter Stopword*.

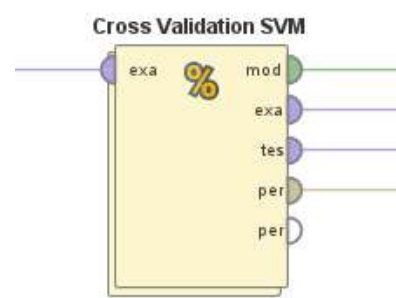
### 3.4. Pengujian Data

Dalam pengumpulan Data terkait analisis sentimen setelah dilakukan *preprocessing* sangat mungkin terjadi *imbalance* data. *Imbalance* data merupakan kondisi dimana di antara kedua kelas memiliki selisih yang sangat signifikan. Pemodelan dengan algoritma yang tidak memperhatikan ketidakseimbangan data akan didominasi oleh kelas mayor dan tidak memperhatikan kelas minornya. Proses menyeimbangkan jumlah kelas dari *dataset* menggunakan Metode *SMOTE*.

Proses Pengujian ini menggunakan menggunakan algoritma SVM dan NB untuk melakukan analisis sentimen dataset

akan dilakukan pengujian dengan menggunakan metode *Fold Cross Validation*. *Fold cross validation* merupakan metode statistik yang digunakan untuk mengevaluasi kinerja model atau algoritma yang akan memisahkan data menjadi dua *subset* yaitu sebagai data proses pembelajaran (data *training*) dan data validasi (data *testing*) sebagaimana disajikan pada Gambar 5.

Dengan metode 10 *Fold Cross validation*, dataset dibagi menjadi 10 area, dengan masing-masing aspek memberikan informasi yang sama persentase setiap jenis data. 9/10 area data digunakan dalam proses *Training* sehingga membentuk model, sedangkan yang 1/10 area digunakan dalam proses *Testing*. *Training* untuk menghasilkan model dan *testing performance*. sebagaimana disajikan pada Gambar 6 dan 7.



**Gambar 5. K-Folds Cross Validation**



**Gambar 6. Training dan Testing Support Vector Machine**





**Gambar 7. Training dan Testing Naive Bayes**

selanjutnya yaitu melakukan tahap evaluasi atas hasil pengujian data. Evaluasi dilakukan dengan menggunakan nilai akurasi pada tabel *confusion matrix* dan nilai *AUC* pada kurva *ROC*

### 3.5. Evaluasi Hasil

Setelah melalui proses tahapan pengujian data, berikut ini merupakan evaluasi hasil dari masing-masing algoritma, yaitu *Support Vector Machine* dan *Naive Bayes*:

#### a. Support Vector Machine

*Confusion matrix* yang dihasilkan dari algoritma SVM dengan Model *SMOTE* sebagaimana disajikan pada tabel 1.

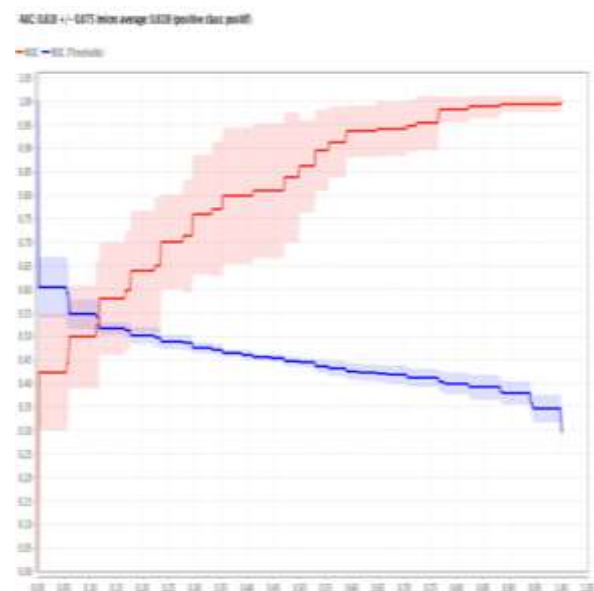
**Tabel 1. Confusion Matrix Support Vector Machine**

accuracy: 72,57% +/- 5,12% (micro average: 72,54%)			
	<i>true Negatif</i>	<i>true Positif</i>	<i>class precision</i>
pred. Negatif	145	67	68,40%
pred. Positif	28	106	76,10%
<i>class recall</i>	83,82%	61,27%	

Pada tabel 1 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 72,57% dengan toleransi kesalahan sebesar 5,12%, dengan nilai *true positif* 106 *records* dan *true negatif*

145 *records*. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai dengan sentimen positif sebanyak 106 *records* dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 145 *records*. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan negatif (*false negatif*) sebanyak 28 *records* dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan positif (*false positif*) sebanyak 67 *records*.

Berdasarkan hasil pengujian *performance* menghasilkan Kurva *Receiver Operating Characteristic* (ROC) sebagaimana pada Gambar 8 dan nilai *Area Under Curve* (AUC) yang didapatkan adalah sebesar 0,810 (*good classification*).



**Gambar 8. Grafik ROC Support Vector Machine**

### b. Naïve Bayes

*Confusion matrix* yang dihasilkan dari algoritma NB dengan model *SMOTE*. Sebagaimana disajikan pada tabel 2.

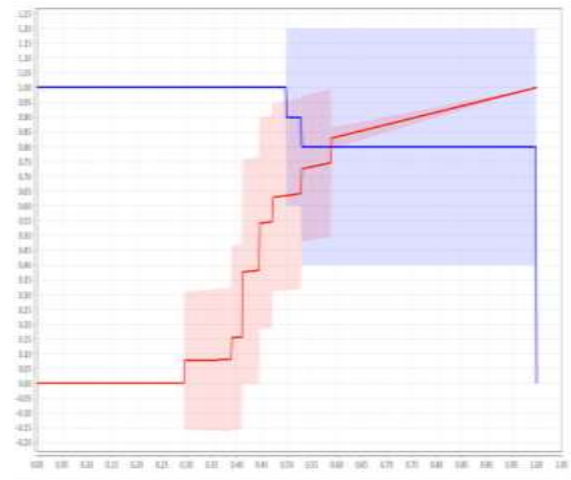
**Tabel 2.** *Confusion Matrix Naïve Bayes*

accuracy: 66,45% +/- 3,66% (micro average: 66,47%)			
	<i>true Negatif</i>	<i>true Positif</i>	<i>class precision</i>
pred. Negatif	95	38	71,43%
pred. Positif	78	135	63,38%
<i>class recall</i>	54,91%	78,03%	

Pada Tabel 2 tersebut menunjukkan bahwa nilai akurasi yang didapatkan sebesar 66,45% dengan toleransi kesalahan sebesar 3,66%, dengan nilai *true positif* 135 *records* dan *true negatif* 95 *records*. Hal ini menunjukkan bahwa data yang diklasifikasikan sesuai dengan sentimen positif sebanyak 135 *records* dan data yang diklasifikasikan sesuai dengan sentimen negatif sebanyak 95 *records*. Sedangkan data yang seharusnya sentimen positif namun diklasifikasikan negatif (*false negatif*) sebanyak 78 *records* dan data yang seharusnya diklasifikasikan sentimen negatif namun diklasifikasikan positif (*false positif*) sebanyak 38 *records*.

Berdasarkan hasil pengujian *performance* menghasilkan Kurva *Receiver Operating Characteristic* (ROC) sebagaimana pada Gambar 9 dan nilai *Area Under Curve* (AUC) yang

didapatkan adalah sebesar 0,499 (*poor classification*).



**Gambar 9.** Grafik ROC Naive Bayes

Berdasarkan analisis rekapitulasi tersebut menunjukkan bahwa evaluasi menggunakan algoritma SVM mendapatkan nilai *accuracy* tertinggi dibandingkan dengan algoritma NB. Berdasarkan hasil penelitian tersebut metode SVM mengungguli metode NB, Hasil penelitian menunjukkan nilai akurasi untuk algoritma NB sebesar 66,45%, Sedangkan nilai akurasi algoritma SVM lebih besar dari NB yaitu sebesar 72,57%. Penelitian ini berhasil mendapatkan algoritma SVM sebagai algoritma yang efektif dan terbaik dalam mengklasifikasikan komentar positif dan komentar negatif terkait dengan sentimen analisis terhadap Kementerian kesehatan.

## IV. KESIMPULAN

Dari penelitian yang telah dilakukan maka dapat diambil kesimpulan bahwa

analisis sentimen dapat digunakan untuk mengukur sentimen masyarakat khususnya media sosial *twitter* terhadap Menteri kesehatan. Tujuannya penelitian ini adalah membantu masyarakat menentukan berapa banyak sentimen yang terdapat pada *tweet* opini Bahasa Indonesia yang ada di *Twitter* yang ditujukan kepada Kementerian Kesehatan. Dalam penelitian mengenai analisis sentimen terhadap Menteri Kesehatan diketahui algoritma *Support Vector Machine* memiliki akurasi lebih tinggi mengungguli algoritma *Naïve Bayes*, besar harapan hasil penelitian ini dapat bermanfaat dan membantu pemerintah dalam mendukung pengambilan keputusan selanjutnya dalam menentukan program Pemerintah dalam memutuskan rantai Covid-19. Selain kesimpulan di atas, penelitian ini juga berupaya memberikan saran untuk penelitian lebih lanjut, terkait dengan jumlah data, penelitian berikutnya dapat meningkatkan jumlah dataset sebagai penelitian karena semakin banyak jumlah datasetnya semakin baik hasilnya melakukan pengujian dengan menggunakan algoritma yang berbeda dari penelitian ini.

#### DAFTAR PUSTAKA

- [1] M. Kamyab, R. Tao, M. H. Mohammadi, and A. Rasool, "Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews," *ACM Int. Conf. Proceeding Ser.*, pp. 14–19, 2018, doi: 10.1145/3293663.3293687.
- [2] J. Weng, E.-P. Lim, J. Jiang, and Z. Qi, *Twitterrank: Finding Topic-Sensitive Influential Twitterers*. 2010.
- [3] J. GoldeNBERG and M. Fresko, "Mine Your Own Business : Market-Structure Surveillance Through Text Mining," vol. 31, no. 3, pp. 521–543, 2012.
- [4] B. Liu, "Sentiment Analysis and Subjectivity," pp. 1–38, 2010, doi: 10.1016/j.dss.2012.05.022.
- [5] R. Jannati, R. Mahendra, C. W. Wardhana, and M. Adriani, "Stance Classification Towards Political Figures on Blog Writing Stance Classification towards Political Figures on Blog Writing," no. February 2019, 2018, doi: 10.1109/IALP.2018.8629144.
- [6] B. Pratama, D. D. Saputra, D. Novianti, and E. P. Purnamasari, "Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM And NB Methods Sentiment Analysis of the

- Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM And NB Methods,” 2019, doi: 10.1088/1742-6596/1201/1/012038.
- [7] A. Tripathy, A. Agrawal, and S. K. Rath, “Classification of Sentimental Reviews Using Machine Learning Techniques,” *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015, doi: 10.1016/j.procs.2015.07.523.
- [8] S. Kurniawan, W. Gata, D. A. Puspitawati, M. Tabrani, and K. Novel, “Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada,” vol. 1, no. 10, pp. 2–8, 2021.
- [9] A. Fauzi, A. N. Rais, M. F. Akbar, and W. Gata, “ANALISIS SENTIMEN BERINTERNET PADA MEDIA SOSIAL NAIVE BAYES,” pp. 46–54, 2018.
- [10] N. Friedman, M. Linial, and D. Pe, “Using Bayesian Networks to Analyze Expression Data,” pp. 127–135, 2000, doi: DOI:10.1089/10665270075005096 1.
- [11] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, “Bayesian Network based Extreme Learning Machine for Subjectivity Detection,” *J. Franklin Inst.*, 2017, doi: 10.1016/j.jfranklin.2017.06.007.
- [12] C. Series, “K-nearest neighbor analysis to predict the accuracy of product delivery using administration of raw material model in the cosmetic industry ( PT K-nearest neighbor analysis to predict the accuracy of product delivery using administration of raw material model in the cosmetic industry ( PT Cedefindo ),” 2019, doi: 10.1088/1742-6596/1367/1/012008.
- [13] L. Dey, “Sentiment Analysis of Review Datasets using Naïve Bayes ’ and K -NN Classifier,” vol. 8, no. 4, p. pp.54-62, 2016, doi: 10.5815/ijieeb.2016.04.07.
- [14] N. D.; Putranti and E. Winarko, “Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine,” vol. 8, no. 1, pp. 91–100, 2014.
- [15] G. Mann, R. Mcdonald, and N. Silberman, “Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models,” pp. 1–9, 2009.
- [16] A. Severyn, “Twitter Sentiment Analysis with Deep Convolutional

- Neural Networks,” pp. 959–962, 2015, doi: 10.1007/978-3-319-19644-2\_60.
- [17] M. G. Santos, Cícero Nogueira dos, “Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts,” pp. 69–78, 2014.
- [18] H. A. Putranto, O. Setyawati, and A. L. Belakang, “Pengaruh Phrase Detection dengan POS -Tagger terhadap Akurasi Klasifikasi Sentimen menggunakan SVM,” vol. 5, no. 4, pp. 252–259, 2016.
- [19] J. Watori, A. Bayhaqy, and H. Pratama, “Sistem Rekomendasi Untuk Pengguna E-Commerce,” vol. 2018, no. Sentika, pp. 23–24, 2018.
- [20] W. Gata and A. Bayhaqy, “Analysis sentiment about islamophobia when Christchurch attack on social media,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 4, pp. 1819–1827, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14179.
- [21] A. S. Nugroho, A. B. Witarto, and D. Handoko, “Support Vector Machine,” 2003.
- [22] J. Chen, H. Huang, S. Tian, and Y. Qu, “Feature Selection for Text Classification with Naive Bayes,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, Apr. 2009, doi: 10.1016/j.eswa.2008.06.054.
- [23] Q. Ye, Z. Zhang, and R. Law, “Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6527–6535, 2009, doi: 10.1016/j.eswa.2008.07.035.
- [24] Alberto Fernandez, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018.
- [25] Sabita, H., & Herwanto, R. (2020). Pantauan Prediktif Covid-19 Dengan Menggunakan Metode SIR dan Model Statistik Di Indonesia. *TEKNIKA*, 14(2), 145-150.