

## ANALISIS SENTIMEN DALAM PENANGANAN COVID-19 DI INDONESIA MENGGUNAKAN *NAIVE BAYES CLASSIFIER*

Tanti Yulianita<sup>1</sup>, Tiani Wahyu Utami<sup>2</sup>, M Al Haris<sup>3</sup>  
*Universitas Muhammadiyah Semarang*

*tantiyulianita1@gmail.com*<sup>1</sup>, *tianiutami@unimus.ac.id*<sup>2</sup>, *alharis@unimus.ac.id*<sup>3</sup>

**Abstrak.** Kasus *Coronavirus Disease* (covid-19) di Indonesia telah berdampak dalam segala lapisan kehidupan. Salah satu kebijakan yang dilakukan oleh pemerintah menjadi sorotan di media sosial yaitu tentang adanya kebijakan Pembatasan Sosial Berskala Besar (PSBB). Banyaknya tanggapan masyarakat tentang kebijakan tersebut sangat beragam terutama di media sosial *Twitter*. Penelitian ini bertujuan mengetahui bagaimana sentimen masyarakat terhadap kebijakan PSBB melalui tanggapan di media sosial *twitter*. Data yang digunakan dengan rentang waktu April – Juni 2020. Data tersebut diklasifikasikan menggunakan algoritma *Naïve Bayes Classifier*. Hasil akurasi yang didapatkan dengan menggunakan *Confusion Matrix* untuk algoritma *Naïve Bayes Classifier* sebesar 89.13%. Sedangkan peluang kesalahan klasifikasi yang dihasilkan oleh kedua metode tersebut dengan menggunakan APER (*Apparent Error Rate*) dengan hasil *Naïve Bayes Classifier* sebesar 10.87%.

**Kata Kunci:** *Analisis Setimen, Covid-19, Naive Bayes Classifier, PSBB, Twitter*

## 1. Pendahuluan

Coronavirus Disease (Covid-19) merupakan salah satu topik yang masih hangat diperbincangkan sejak awal kemunculannya di akhir tahun 2019 sampai saat ini. Berdasarkan data dari *World Health Organization* (WHO) 30 Juni 2020 sebanyak 216 negara kasus yang terkonfirmasi kasus covid-19 (WHO, 2020). Indonesia merupakan salah satu negara yang terkonfirmasi terpaparnya covid-19 yang masuk pada awal maret 2020. Menurut data yang dikeluarkan oleh Gugus Percepatan Penanganan Covid-19 pada 30 Juni 2020 terdapat 56.325 yang positif terpapar virus tersebut, data covid-19 di Indonesia masih bersifat fluktuatif. Awal kemunculan covid-19 di Indonesia, pemerintah telah melakukan berbagai upaya kebijakan yang timbul akibat virus tersebut mulai dari bidang kesehatan, sosial, ekonomi dan bidang lainnya.

Seiring dengan kebijakan pemerintah yang sedang dilakukan saat ini tidak luput dari komentar publik dalam media sosial *twitter* khususnya yang paling disoroti yaitu dengan adanya pemberlakuan Pembatasan Sosial Berskala Besar (PSBB). Kebijakan yang menuai berbagai macam komentar, mulai dari pujian, kritik, saran bahkan sindiran. Berdasarkan hasil survei Indo Barometer yang diberitakan melalui laman *kompas.com* disebutkan sebanyak 40.3% responden yang menolak PSBB dengan alasan kesulitan mencari nafkah. Secara keseluruhan, hasil survey tersebut menunjukkan bahwa 10% respon yang menganggap bahwa penerapan PSBB merupakan langkah yang tidak tepat, sementara 89.5% responden menilai bahwa penerapan PSBB sudah sanga tepat (Purnamasari, 2020). Oleh karena itu, dapat dibuat suatu pendekatan untuk mengetahui persepsi publik terhadap kebijakan pemerintah khususnya PSBB. Banyaknya pengguna *twitter* yang menyampaikan berbagai tanggapan tersebut dapat dimanfaatkan dengan menganalisis sentimen terhadap kebijakan pemerintah dalam penanganan covid-19 khususnya PSBB.

Analisis sentimen atau bisa disebut juga dengan Tanggapan Sentimen yang tujuannya untuk menganalisis, mengolah, mengekstraksi data tekstual yang berupa tanggapan terhadap suatu entitas terhadap suatu topik guna memperoleh sebuah informasi (Rozi, Pramono, & Dahlan, 2012). Analisis sentimen dilakukan untuk

melihat pendapat atau kecenderungan tanggapan terhadap sebuah masalah apakah cenderung berpandangan atau bertanggapan negatif atau positif dengan berbagai macam metode pendekatan.

Terdapat beberapa penelitian yang sudah pernah dilakukan berkaitan dengan analisis sentimen dengan pendekatan klasifikasi. Penelitian yang dilakukan oleh Gusriani, Kartina, & Muhammad (2016) dengan judul artikel “Analisis Sentimen Terhadap Toko Online Di Sosial Media Menggunakan Metode Klasifikasi *Naïve Bayes Classifier* (Studi Kasus: Facebook Page BerryBenka)”. Dari hasil penelitiannya menunjukkan bahwa NBC dapat dijadikan metode klasifikasi untuk analisis sentimen dengan keakuratan 93.7%. Selanjutnya penelitian lain dilakukan oleh Mahardika & Eri (2018), yakni penggunaan *Naïve Bayes Classifier* dalam melakukan klasifikasi yang berjudul “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes Classifier*”. Tujuan dari penelitian tersebut untuk mengklasifikasikan sentimen ke dalam kelas negatif dan positif dengan 300 data latih dan 100 data uji untuk mendapatkan tingkat keakuratan 97%. Menurut Muel, Utami & Rochdi (2019) tentang “Analisis Sentimen Pada *Twitter* Gojek dengan Metode *Naive Bayes Classifier* Menggunakan Visualisasi *Latent Dirichket Allocation*” dengan hasil akurasi metode *Naive Bayes Classifier* sebesar 78.55%.

## 2. Metode Penelitian

### 2.1 Sumber dan Struktur Data

Data yang digunakan adalah data sekunder. Data diambil dengan menggunakan *Python*. Populasi pada penelitian ini adalah seluruh komentar menggunakan tagar “#PSBB” di twitter. Sampel yang digunakan adalah data diambil sebanyak satu bulan yaitu April - Juni 2020.

Variabel yang digunakan sebanyak dua atribut dan satu label. Atribut yang digunakan antara lain tanggal *Tweet* dan *Tweet*. Label yang digunakan untuk pengklasifikasian adalah data hasil pelabelan dari analisis sentimen. Jenis data label adalah kategorik yakni positif, dan negatif.

Tabel 1. Atribut Penelitian

No	Indikator	Jenis Data	Keterangan
1	Tanggal <i>Tweet</i>	Date	Atribut
2	Kicauan ( <i>Tweet</i> )	String	Atribut

Atribut tanggal *Tweet* berisi data mulai dari tanggal April - Juni 2020. *Tweet* berisi data tanggapan pengguna twitter mengenai #PSBB, komentar dapat berupa dalam Bahasa Indonesia, singkatan, slang (bahasa gaul).

Penulis akan menilai sentimen yang terkandung di dalam *Tweet* tersebut dan menandai *Tweet* tersebut ke dalam 2 kategori sentimen yaitu *Tweet* yang mengandung sentimen negatif dan positif.

## 2.2 Langkah Penelitian

Penelitian ini menggunakan aplikasi *Python*, *RStudio*, dan *Rapidminer*. Langkah penelitian ini adalah sebagai berikut:

- a. *Crawling data* (mengambil data) *Tweet* dengan *Python*
  - 1) Memasukkan *keyword* #PSBB
  - 2) Menyimpan data yang diperoleh
- b. Menyiapkan data *Tweet*, kamus *stopwords*, kamus *slangword* dan kamus *stemming*
- c. Tahapan *Preprocessing Text* dengan *Rstudio*
  - 1) **Cleansing**, membersihkan *Tweet* yang tidak diperlukan untuk mengurangi noise. Penghilangan karakter HTML, kata kunci, ikon emosi, *hashtag* (#), *username*, url (<http://situs.com>) ataupun e-mail (Aditya, 2015).
  - 2) **Case Folding**, merubah setiap kaat menjadi sama misalnya huruf besar (*uppercase*) didalam *Tweet* menjadi huruf kecil (*lowercase*) (Salam, Junta, & Rima, 2018)..
  - 3) **Tokenizing**, memenggal setip kata pada suatu kalimat dan pada saat bersamaan membuang karakter tertentu yang dianggap sebagai tanda baca (Asiyah, 2016).
  - 4) **Stopword**, pengecekan setiap kata pada suatu *tweet*, selanjutnya kata-kata yang terlalu umum dan kurang penting.
  - 5) **Stemming**, proses mendapatkan kata dasar pada sutau kalimat degan cata menghilangkan awalan, akhiran, sisipan, atau kombinasi ketiganya.

- d. Pelabelan, setelah tahapan *pre processing text*, memberi label pada *Tweet* dengan label positif dan negatif.
- e. Membagi data sebagai *data training* dan *data testing*.
- f. *Naïve Bayes Classifier*

1) Menghitung probabilitas dari  $v_j$  pada data *training* dengan persamaan

$$P(a_i|v_j) = \frac{P(v_j|a_i) P(a_i)}{P(v_j)}$$

dimana  $v_j$  merupakan kategori sentimen, yaitu

$v_1$  = positif, dan  $v_2$  = negatif.

2) Menghitung probabilitas kata  $a_i$  pada kategori  $v_j$  dengan persamaan

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

3) Model probabilitas NBC disimpan dan digunakan untuk tahap data *testing*.

4) Menghitung probabilitas tertinggi dari kategori sentimen yang diujikan

$$V_{MAP} \text{ dengan persamaan } V_{MAP} = \underset{A_j}{\operatorname{argmax}} P(v_j) \prod_{i=1}^n P(a_i|v_j)$$

5) Mencari nilai  $V_{MAP}$  paling maksimum dan memasukkan *tweet* tersebut pada kategori dengan  $V_{MAP}$  maksimum.

### 3. Hasil dan Pembahasan

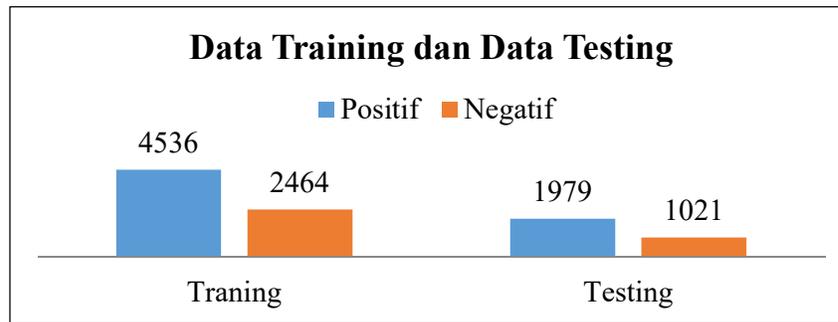
Gambaran umum mengenai data tanggapan publik tentang kebijakan pemerintah dalam penanganan covid-19 di Indonesia khususnya Penerapan Pembatasan Sosial Berskala Besar (PSBB) yang telah diperoleh. Data tanggapan terdiri dari data tanggapan pada bulan April – Juni 2020 yang dikategorikan kedalam dua kelas kategori, yaitu tanggapan positif dan tanggapan negatif. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori ditunjukkan Gambar 1.



**Gambar 1.** Persentase Tanggapan Publik Terhadap PSBB

Berdasarkan Gambar 1 diketahui bahwa dari total tanggapan publik terhadap kebijakan pemerintah dalam penanganan covid-19 khususnya dengan adanya

penerapan pembatasan sosial berskala besar (PSBB) di Indonesia yang diperoleh yaitu 10000 tanggapan, sebanyak 65% atau sebanyak 6515 tanggapan merupakan tanggapan positif yaitu tanggapan yang mendukung diterapkannya PSBB sedangkan 35% atau sebanyak 3485 tanggapan merupakan tanggapan negatif yaitu tanggapan yang kurang setuju dengan adanya sistem PSBB tersebut.



**Gambar 2.** Data *Training* dan Data *Testing*

Dalam Gambar 2. merupakan pembagian data *training* dan data *testing* yang sudah melalui proses *pre-processing text* dengan pembagian 70% sebagai data *training* dan 30% data *testing*. Dalam data *training* sebanyak 4536 masyarakat beranggapan positif tentang penerapan PSBB sedangkan 2464 beranggapan negatif. Sedangkan untuk data *testing* sebanyak 1979 beranggapan positif dan 1021 beranggapan negatif.

### 3.1 Klasifikasi *Naïve Bayes Classifier*

Pada tahap ini dilakukan evaluasi pada klasifikasi menggunakan *Naïve Bayes Classifier* dan *K-Nearest Neighbour* dengan pembagian data *training* sebesar 70% dan data *testing* sebesar 30%. Hasil pertama diperoleh persamaan pada *confusion matrix* klasifikasi pada setiap kelas. Berikut hasil yang didapatkan:

**Tabel 2.** *Naïve Bayes Classifier*

Aktual	Prediksi	
	Positif	Negatif
Positif	1979	0
Negatif	326	695
<i>Accuracy</i>	89.13%	

Berdasarkan hasil perhitungan menggunakan *confusion matrix* pada algoritma *Naïve Bayes Classifier* data *Tweets* diatas prediksi benar sebanyak 2674 dan

prediksi salah sebanyak 695. Hasil klasifikasi dengan algoritma *Naïve Bayes Classifier* untuk PSBB diperoleh nilai *accuracy* sebesar 89.13%.

**Tabel 3.** Evaluasi Performa

<b>Klasifikasi</b>	<b><i>Accuracy</i></b>	<b><i>Recall</i></b>	<b><i>F-Measure</i></b>
<b><i>Naïve Bayes Classifier</i></b>	89.13%	100%	92.39%

Berdasarkan evaluasi performa yang disajikan pada Tabel 3 untuk algoritma *Naïve Bayes Classifier* dengan tiga evaluasi *precision*, *Recall*, *F-Measure*. Algoritma *Naïve Bayes Classifier* dengan *Recall* sebesar 100%, dan *F-Measure* sebesar 92.39%.

**Tabel 4.** *Apparent Error Rate*

<b>Klasifikasi</b>	<b>APER</b>
<b><i>Naïve Bayes Classifier</i></b>	10.87%

Berdasarkan Tabel 4 peluang kesalahan klasifikasi yang dihasilkan oleh *Naïve Bayes Classifier* sebesar 10.87%. Semakin kecil nilai APER maka hasil pengklasifikasian semakin baik.

#### 4. Kesimpulan

Berdasarkan pembahasan pada pembahasan sebelumnya maka dapat ditarik beberapa kesimpulan, sebagai berikut:

- a. Hasil deskripsi tanggapan mengenai Pembatasan Sosial Berskala Besar (PSBB) dalam penanganan covid-19 di Indonesia dengan jumlah *Tweet* atau tanggapan sebanyak 10000 *Tweet* dari Bulan April- Juni 2020. Dalam *Tweet* tersebut terdapat 6.515 *Tweet* yang mengandung sentimen positif sedangkan 3.485 *Tweet* mengandung sentimen negatif.
- b. Hasil klasifikasi sentimen *twitter* mengenai PSBB dalam penanganan covid-19 di Indonesia dengan menggunakan *Naïve Bayes Classifier* didapatkan jumlah sentimen positif sebanyak 2.305 dan sentimen negatif sebanyak 695.
- c. Hasil akurasi yang didapatkan dengan menggunakan *Confussion Matriksi* untuk algoritma *Naïve Bayes Classifier* sebesar 89.13%. Untuk melihat peluang kesalahan klasifikasi yang dihasilkan oleh kedua metode tersebut dengan menggunakan APER (*Apparent Error Rate*) dengan hasil *Naïve Bayes Classifier* sebesar 10.87%. Semakin kecil nilai APER maka hasil

pengklasifikasian semakin baik. Hal ini menunjukkan kemampuan prediksi dengan menggunakan *Naïve Bayes Classifier* baik sehingga dapat dijadikan sebagai rujukan untuk mengolah data *text*.

### **5. Ucapan Terima Kasih**

Terima kasih kepada semua pihak yang berperan dalam penelitian ini terkhusus Ibu Tiani Wahyu Utami, M.Si dan Bapak M Al Haris selaku Dosen Pembimbing yang support untuk mengikuti seminar nasional ini. Terima kasih kepada Orang tua saya yang sudah memberi doa dan supportnya.

## Daftar Pustaka

- Aditya, B. R. (2015). Penggunaan Web Crawler untuk Menghimpun *Tweets* dengan Metode Pre-Processing Text Mining. *Infotel*, 7(2), 93-100.
- Asiyah, S. N. (2016). Klasifikasi Berita *Online* Menggunakan Metode *Support Vector Machine* dan *K-Nearest Neighbour*. Surabaya: Institut Teknologi Sepuluh November.
- Gugus COVID-19. (2020). Homepage Gugus Tugas Percepatan Penangan COVID-19. <https://covid19.go.id/> . Tanggal 30 Juni 2020 Pukul 09.00.
- Gusriani, S., Kartina, D. K. W., & Muhammad, I.Z. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka). Pekanbaru: Universitas Caltex Riau.
- Mahardika, Y. S., & Eri, Z. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial *Twitter* Menggunakan Algoritma *Naive Bayes Classifier* . *Prosiding Sintax*. Universitas Stikubank.
- Muel, S.S., Utami, T.W., & Rochdi, W. (2019). Analisis Sentien Pada *Twitter* Dengan Metode *Naive Bayes Classifier* Menggunakan Visualisasi *Latent Dirichilet Allocation*. Semarang: Universitas Muhammadiyah Semarang.
- Purnamasari, D.M. (2020). Survei: Sebagian Besar Warga Menolak PSBB karena Sulit Cari Nafkah. <https://nasional.kompas.com/read/2020/04/23/13541671/survei-sebagian-besar-warga-menolak-psbb-karena-sulit-cari-nafkah>. Diakses 23 April 2020 (09.30)
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi *Tanggapanon Mining* (Analisis Sentimen) untuk Ekstaksi Data Tanggapan Publik pada Perguruan Tinggi. *EECCIS*, Juni, 6(1), pp. 37-43.
- Salam, A., Junta, Z., & Rima, S. P. U. K. (2018). Analisis Sentimen Data Komentar Sosial Media *Facebook* Dengan *K-Nearest Neighbour* (Studi Kasus Pada Akun Jasa Ekspedisi Barang JdanT Ekspre Indonesia). *Prosiding Sintax*. Universitas Dian Nuswantoro.
- World Health Organization. (2020). Coronavirus disease (COVID-19) [https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=Cj0KCQjww\\_f2BRCARIsAP3zarGHHhl2cj59Tg1uPCU9Hf3jNXEYcow81FnuNA3rX1joQ0GjhRLhSxgaAt07EALw\\_wcB](https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=Cj0KCQjww_f2BRCARIsAP3zarGHHhl2cj59Tg1uPCU9Hf3jNXEYcow81FnuNA3rX1joQ0GjhRLhSxgaAt07EALw_wcB) . Diakses 9 Juni 2020 (08:30)