

## Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter

Muhammad Dwison Alizah<sup>1</sup>, Arifin Nugroho<sup>2</sup>, Ummu Radiyah<sup>3</sup>, Windu Gata<sup>4</sup>

<sup>1,2,3,4</sup>Prodi Ilmu Komputer, STMIK Nusa Mandiri

<sup>1</sup>14002299@nusamandiri.ac.id, <sup>2</sup>14002306@nusamandiri.ac.id, <sup>3</sup>ummu.urd@nusamandiri.ac.id,

<sup>4</sup>windu@nusamandiri.ac.id

**Abstract:** Covid-19 has been set as a Pandemic by the World Health Organization (WHO). The very large impact and the infection that is fast enough are the reasons for making Covid-19 as a pandemic and efforts to overcome. One anticipation that can be done is to do lockdown. Making the decision to carry out a lockdown is intended to reduce the spread that occurs. Lockdown is certainly not a 100% good solution for all of individual. There are individual who agree that the lockdown will be implemented, also there are those who think that the lockdown is better not to be carried out considering the negative impacts that can occur. Therefore in this study will be presented the predictive modeling for sentiment analysis related to "lockdown" specially on social media Twitter. The method used to labeled was using Vader then the tweets are extracted using TF-IDF, and modeling is made for the prediction of sentiment using Naïve Bayes and Support Vector Machine. The results obtained from the two algorithms are more than 80%.

**Keywords:** Covid-19, lockdown, TF-IDF, Naïve Bayes, Support Vector Machine

**Abstrak:** Covid-19 telah ditetapkan sebagai Pandemi oleh World Health Organization (WHO). Dampak yang sangat besar dan penyebaran yang cukup cepat menjadi alasan untuk menjadikan Covid-19 sebagai Pandemi dan perlu dilakukan upaya penanggulangan. Salah satu upaya yang bisa dilakukan adalah dengan melakukan lockdown. Pengambilan keputusan untuk melakukan lockdown diperuntukan guna mengurangi penyebaran yang terjadi. Lockdown tentunya bukanlah solusi yang 100% baik bagi segala pihak. Terdapat pihak - pihak yang menyetujui akan dilaksanakannya lockdown, ada pula yang beranggapan bahwa lockdown lebih baik tidak dilaksanakan dengan pertimbangan dampak negatif yang bisa terjadi. Oleh karena itu, pada penelitian ini akan disampaikan mengenai pembuatan pemodelan prediksi terkait analisa sentimen terkait "Lockdown" yang dikhususkan pada media sosial Twitter. Metode yang digunakan adalah dengan melakukan labeling menggunakan Vader dan selanjutnya tweet tersebut dilakukan ekstraksi menggunakan TF-IDF, dan dibuatkan pemodelan untuk prediksi sentimen menggunakan Naïve Bayes dan Support Vector Machine. Hasil evaluasi yang didapat dari kedua algoritma tersebut ialah mencapai lebih dari 80%.

**Kata kunci:** Covid-19, lockdown, TF-IDF, Naïve Bayes, Support Vector Machine



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2019 by author and IJSE-Indonesian Journal on Software Engineering.

### A. PENDAHULUAN

Novel coronavirus 2019 (nCoV-2019), atau yang sekarang disebut dengan *corona virus disease-19* (COVID-19) sudah ditetapkan sebagai Pandemi oleh *World Health Organization* (WHO) (Cucinotta & Vanelli, 2020). Hal ini berakibat banyak negara mulai melakukan "lockdown" untuk mengantisipasi penyebaran virus ini (Lau et al., 2020). Di Indonesia, hal ini belum dilakukan oleh Pemerintah walaupun terdapat beberapa penduduk Indonesia menyuarkan "lockdown" demi menekan penyebaran virus. Tentunya banyak hal yang perlu dipertimbangkan pemerintah tentang dampak dari penerapan "lockdown" bagi Indonesia.

Twitter adalah penyedia layanan pesan yang menyediakan begitu banyak karakteristik dengan alat komunikasi yang digunakan ("Text Mining Dan Sentimen Analisis Twitter Pada Gerakan Lgbt," 2017). Twitter termasuk salah satu media sosial yang ramai digunakan untuk

berbagi informasi. Salah satu faktornya ialah penggunaannya yang mudah dan mobilitas yang tinggi. Pertukaran informasi pada Twitter termasuk dalam kategori yang sangat cepat. Pengguna bisa berbagi informasi apa saja dan menautkan tagar untuk mempermudah pencarian informasi.

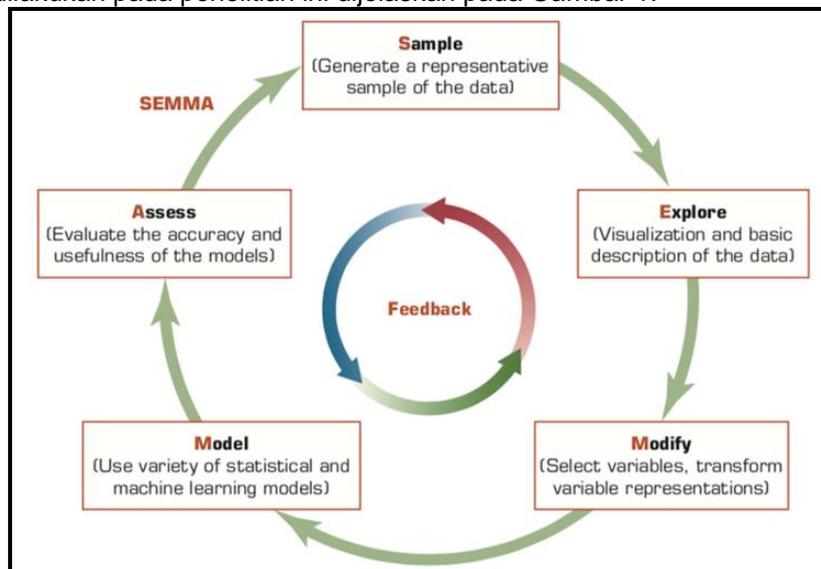
Pengguna Twitter beberapa hari kebelakang banyak sekali yang berbagi informasi mengenai Covid-19 setelah virus ini ditetapkan sebagai Pandemi dan telah masuk di Indonesia. Tidak sedikit pula yang menyuarakan untuk melakukan *lockdown* melalui Twitter. Hal inilah yang mendorong penelitian ini untuk membuat sebuah pemodelan analisis sentimen terkait dengan *tweets* mengenai “*lockdown*” di Indonesia.

Skema *lockdown* dilakukan di beberapa Negara guna menekan angka penyebaran pandemi Covid-19. Tentunya banyak Pro dan Kontra terkait dengan pemberlakuan skema tersebut. Oleh karena itu, penulis mencoba untuk menganalisa dan membuat pemodelan *machine learning* terkait dengan analisis sentimen dari *tweets* “*lockdown*”. Analisis sentimen merupakan suatu proses pengolahan informasi untuk mengelompokan atau mengklasifikasi opini, penilaian seseorang terkait kreasi, organisasi, atau aktivitas tertentu (Buntoro et al., 2019). Tujuan dari melakukan analisa sentimen ialah untuk menentukan perilaku dan opini penulis terhadap topik tertentu (Pramono, Didi Rosiyadi, & Windu Gata, 2019).

Pemodelan yang akan dilakukan pada penelitian ini menggunakan metode *Support Vector Classification* dan *Naïve-Bayes*. *Support Vector Machine* (SVM) dan *Naïve-Bayes* merupakan metode yang bisa digunakan baik untuk klasifikasi maupun regresi (Bustami, 2014; Ma & Guo, 2014), (Frank, Trigg, Holmes, & Witten, 2000). Beberapa penelitian sebelumnya terkait sentimen analisis pada twitter seperti (Buntoro et al., 2019) dan (Pratama & Sarno, 2016) menggunakan *Support Vector Classification* dan *Naïve-Bayes* dan menghasilkan performa yang baik dan cukup bervariasi. Pada penelitian kali ini, SVM dan *Naïve-Bayes* digunakan sebagai metode klasifikasi untuk memprediksi label sentimen pada suatu *tweets* pada twitter.

## B. METODE PENELITIAN

Pada penelitian ini, penulis mencoba menggunakan beberapa metode *Natural Language Processing* (NLP) dan *Machine Learning*. Secara umum metodologi yang dilakukan pada penelitian ini mengacu kepada SEMMA Data Mining Process (Azevedo & Santos, 2008). Tahapan yang dilakukan pada penelitian ini dijelaskan pada Gambar 1.



Gambar 1. Semma Data Mining Process (Wardhani, 2015)

### 1. Sample

Pada tahap Pengumpulan Data, dilakukan pengumpulan data terkait *tweets* “*lockdown*” pada media sosial twitter memanfaatkan API Twitter. Data yang dikumpulkan disimpan dalam format JSON pada sebuah file. Selanjutnya data tersebut dibaca dan dilakukan proses persiapan data.

## 2. Explore

Pada tahap ini dijelaskan mengenai deskripsi data dan visualisasi data. Deskripsi data akan menjelaskan mengenai apa gambaran besar dari informasi data yang digunakan. Visualisasi data akan memperlihatkan informasi secara visual dari data.

## 3. Modify

Pada tahap ini dilakukan beberapa metode persiapan data diantaranya:

- a. Tokenisasi  
Tokenisasi dilakukan untuk memecah sekumpulan kata (kalimat) menjadi kata yang memiliki arti tertentu. Hal ini bertujuan agar kata perkata dapat diolah lebih lanjut dalam proses – proses selanjutnya.
- b. Pembersihan URL, *Username*, *Hashtag*  
Pembersihan ini dilakukan untuk menghilangkan URL atau tautan link website, Username (i.e. @professor.parno), dan Hashtag (#COVID19) yang ada pada tweets yang ada pada data.
- c. *Stemming*  
*Stemming* merupakan proses yang digunakan untuk mengembalikan kata – kata ke dalam kata dasarnya. Hal ini juga bertujuan untuk membersihkan suatu kata dengan pengejaan yang kurang tepat.
- d. *Stopword Removal*  
Tahap ini merupakan proses untuk melakukan filter terhadap kata – kata umum seperti “the”, “a”, dan lainnya, yang tidak diperlukan saat pemrosesan data. Kata – kata tersebut dihilangkan dari tweets.
- e. *Labeling*  
Tahap ini merupakan tahap dimana hasil dari tahap sebelumnya akan dilakukan perhitungan terhadap polarity dari tweets yang terambil, sehingga dapat menghasilkan label positif maupun negative terkait sentiment dari *tweets* tersebut.

## 4. Model

Pada tahap ini akan dilakukan proses pembagian data menjadi dua, yaitu data uji dan data latih dengan rasio 3:7. Setelah itu, dilakukan proses ekstraksi fitur dari *tweets* yang terambil sehingga selanjutnya *tweets* tersebut dapat dilakukan proses pemodelan *machine learning* menggunakan data latih untuk melakukan prediksi label pada sentimen *tweets* pada data uji.

## 5. Assess

Pada tahap ini akan dilakukan evaluasi terhadap pemodelan yang telah dibuat. Evaluasi yang dimaksud adalah membandingkan hasil yang didapat dari model prediksi terhadap data uji dengan label sentimen pada data uji yang sebelumnya sudah diberi label. Hasil evaluasi yang dilakukan dihitung berdasarkan besaran dari *precision*, *recall*, *f1-score* dan akurasi.

## C. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil dari penelitian yang dilakukan. Pada percobaan yang dilakukan, penulis menggunakan *python* sebagai *tools* dalam menerapkan konsep yang ditawarkan. Hasil dari penelitian ini akan dibagi menjadi beberapa tahapan sesuai dengan metodologi yang digunakan.

### 1. Hasil pada tahap Sample.

Data yang diambil menggunakan Twitter API disimpan ke dalam *file*. Data yang diambil bersifat mentah dan berisikan keseluruhan atribut yang ada pada tweets di twitter seperti *created\_at*, *id*, *text*, *source*, *name*, *description*, hingga *retweeted\_status*. Atribut yang digunakan nantinya adalah *text*, dimana data ini merupakan keseluruhan *tweet* dari data terambil.

### 2. Hasil pada tahap Explore.

Data *tweet* yang diambil merupakan live tweet pada 19 April 2020 pukul 12:42 sampai dengan 15:02. Tweet yang diambil berdasarkan kata kunci “*lockdown*” yang menghasilkan

data sebanyak 15.494 *tweet*. Pada Tabel 1 ditampilkan tujuh (7) hasil ekstraksi *text* dari *tweets* yang terambil.

Tabel 1 Hasil Ekstraksi Text dari Tweets Terambil

Index	Tweets
0	RT @ShobhaBJP: Horrific visuals from the land ...
1	RT @EveningStandard: Lions nap peacefully on S...
2	@ImranKhanPTI @MJALSHRIKA This is the #Hindu #...
3	RT @HLN_BE: Zuid-Korea versoepelt lockdown bee...
4	I was supposed to have my first cum bukkake ne...
5	RT @KChiruTweets: On a Sunday before lockdown....
6	RT @tanwer_m: This is from Jehanabad, Bihar. \...

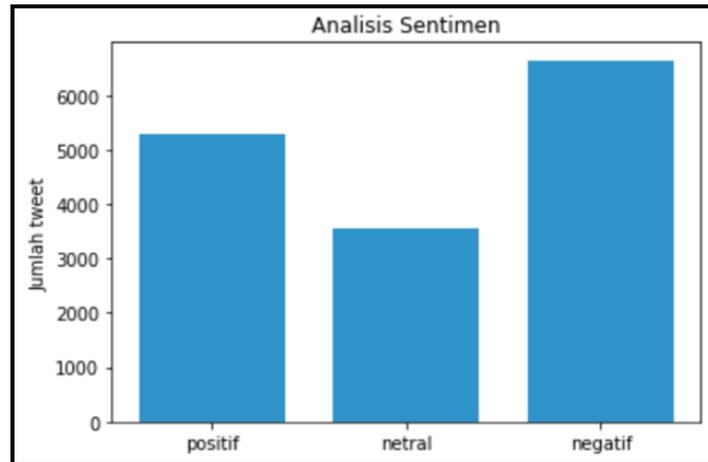
### 3. Hasil pada Modify

*Tweets* yang sudah diambil selanjutnya dilakukan modifikasi untuk nantinya dapat dilakukan proses pembuatan pemodelan machine learning untuk prediksi label sentimen. Pada tahap ini, *tweets* mentah dilakukan proses tokenisasi, pembersihan, *stemming*, *stopword removal* hingga menghasilkan seperti yang ditunjukkan pada tabel 2.

Tabel 2 Hasil Modifikasi Data Tweets

Tweets	Compound	Label
rt horrific visuals land shivajimaharaj hindu ...	-0.7964	negatif
rt lions nap peacefully south africa roads cor...	0.5267	positif
hindu sindhi pakistan ready commit suicide chi...	-0.2023	negatif
rt zuid-korea versoepelt lockdown beetje	0.0000	netral
supposed first cum bukkake next week lockdown ...	-0.4767	negatif
rt sunday lockdown missing meeting dear ones s...	0.8074	positif
rt jehanabad bihar kids surging lockdown eatin...	0.0000	netral

Data yang terlihat pada Tabel 2 menghasilkan data *tweets* hasil dari proses tokenisasi, pembersihan, *stemming*, *stopword removal* dan juga dihasilkan nilai *compound* yang digunakan sebagai angka untuk melakukan pelabelan pada *tweets*.



Gambar 2 Histogram dari Jumlah Sentimen (Positif, Netral, dan Negatif)

#### 4. Hasil pada Proses Model

Pada tahap ini, data yang siap diproses dibagi menjadi dua bagian, yaitu data latih dan data uji dengan rasio 70% untuk data latih dan 30% untuk data uji. Dari total 15.494 data, sebanyak 10.845 dijadikan sebagai data latih, dan 4649 data dijadikan sebagai data uji. Pemodelan machine learning yang digunakan ialah dengan *Naïve Bayes* dan *Support Vector Machine* dengan menggunakan metode ekstraksi fitur yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF).

#### 5. Hasil pada Proses Assess

Evaluasi dari pemodelan yang dibuat dijelaskan menggunakan dua metode yaitu *Confusion Matrix* dan *Classification Report*. Hasil evaluasi tersebut penulis bagi menjadi dua bagian, yaitu sebagai berikut:

##### a. *Naïve Bayes*

Pada Tabel 3 dan tabel 4 berturut-turut adalah hasil evaluasi evaluasi menggunakan *Confusion Matrix* dan *Classification Report* pada percobaan menggunakan algoritma *Naïve Bayes*.

Tabel 3 Confusion Matrix (Naive Bayes)

		Predicted		
		Positif	Netral	Negatif
Actual	Positif	718	105	247
	Netral	52	1655	296
	Negatif	37	142	1397

*Confusion Matrix* pada penerapan *Naïve Bayes* menunjukkan bahwa prediksi benar pada sentiment positif sebanyak 718, prediksi benar pada netral sebanyak 1655, dan prediksi benar terhadap sentiment negatif sebanyak 1397.

Tabel 4 Classification Report (*Naive Bayes*)

	Precision	Recall	F1-score	Support
<b>Negative</b>	0.89	0.67	0.77	1070
<b>Netral</b>	0.87	0.83	0.79	2003
<b>Positif</b>	0.72	0.89	0.79	1576
<b>Accuracy</b>				0.81
<b>Macro avg</b>	0.83	0.79	0.80	4649
<b>Weighted avg</b>	0.82	0.81	0.81	4649

*Classification Report* pada penerapan *Naïve Bayes* menunjukkan hasil evaluasi seperti *precision*, *recall*, *f1-score* dengan rata-rata sebesar 0,80 dan 0,81. .

b. *Support Vector Machine*

Pada Tabel 5 dan tabel 6 berturut-turut adalah hasil evaluasi menggunakan *Confusion Matrix* dan *Classification Report* pada percobaan menggunakan algoritma *Support Vector Machine*.

Tabel 5 Confusion Matrix (SVM)

		Predicted		
		Positif	Netral	Negatif
Actual	Positif	866	109	95
	Netral	60	1835	108
	Negatif	55	164	1357

*Confusion Matrix* pada penerapan SVM menunjukkan bahwa prediksi benar pada sentiment positif sebanyak 886, prediksi benar pada netral sebanyak 1835, dan prediksi benar terhadap sentiment negatif sebanyak 1357.

Tabel 6 Classification Report (SVM)

	Precision	Recall	F1-score	Support
<b>Negative</b>	0.88	0.81	0.84	1070
<b>Netral</b>	0.87	0.92	0.89	2003
<b>Positif</b>	0.87	0.86	0.87	1576
<b>Accuracy</b>				0.87
<b>Macro avg</b>	0.87	0.86	0.87	4649
<b>Weighted avg</b>	0.87	0.87	0.87	4649

*Classification Report* pada penerapan SVM menunjukkan hasil evaluasi seperti *precision*, *recall*, *f1-score* dengan rata-rata sebesar 0,87. .

#### D. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa *Naïve Bayes* dan *Support Vector Machine* memiliki performa yang cukup baik untuk memprediksi sentimen dari suatu *tweets* mengenai *lockdown* pada media sosial Twitter. *Naïve Bayes* mampu mencapai akurasi sebesar 81% dengan *f1-score* 0.8, sedangkan *Support Vector Machine* mampu mencapai akurasi sebesar 87% dengan *f1-score* 0.87.

Saran penelitian lanjutan ialah dengan melakukan penelitian menggunakan data yang lebih banyak dan dengan proses persiapan data yang lebih baik. Juga bisa dilakukan dengan mengganti metode yang digunakan, baik saat ekstraksi fitur, maupun saat pembuatan model prediksi

#### REFERENSI

- Azevedo, A., & Santos, M. F. (2008). KDD, semma and CRISP-DM: A parallel overview. *MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008*.
- Buntoro, G. A., Alvianda, F., Adikara, P. P., Al Kabir, A. H., Yasid, M., Hayuningtyas, R. Y., ... Susanto, A. (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 3(1), 1–10. <https://doi.org/10.22219/repositor.v1i1.11>
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*.
- Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomedica*. <https://doi.org/10.23750/abm.v91i1.9397>

- Frank, E., Trigg, L., Holmes, G., & Witten, I. H. (2000). Naive Bayes for Regression. *Machine Learning*, 41(1), 5–25. Retrieved from <http://www.cs.waikato.ac.nz/~eibe/pubs/nbr.pdf>
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of Travel Medicine*. <https://doi.org/10.1093/jtm/taaa037>
- Ma, Y., & Guo, G. (2014). Support vector machines applications. In *Support Vector Machines Applications*. <https://doi.org/10.1007/978-3-319-02300-7>
- Pramono, F., Didi Rosiyadi, & Windu Gata. (2019). Integrasi N-gram, Information Gain, Particle Swarm Optimization di Naïve Bayes untuk Optimasi Sentimen Google Classroom. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 383–388. <https://doi.org/10.29207/resti.v3i3.1119>
- Pratama, B. Y., & Sarno, R. (2016). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015*. <https://doi.org/10.1109/ICODSE.2015.7436992>
- Text Mining Dan Sentimen Analisis Twitter Pada Gerakan Lgbt. (2017). *Intuisi: Jurnal Psikologi Ilmiah*, 9(1), 18–25.
- Wardhani, P. A. (2015). Business Intelligence, Analytics, and Data Science: A Managerial Perspective. In *Efikasi Diri dan Pemahaman Konsep IPA dengan Hasil Belajar Ilmu Pengetahuan Alam Siswa Sekolah Dasar Negeri Kota Bengkulu* (4th ed., Vol. 6). <https://doi.org/10.1017/CBO9781107415324.004>