

Sentimen Analisis Terkait “Lockdown” pada Sosial Media Twitter

Muhammad Dwison Alizah*¹, Arifin Nugroho², Ummu Radiyah³, Windu Gata⁴

^{1,2,3,4}STMIK Nusa Mandiri; Jakarta, (021) 78839513

e-mail: *¹14002299@nusamandiri.ac.id, ²14002306@nusamandiri.ac.id, ³ummu.urd@nusamandiri.ac.id, ⁴windu@nusamandiri.ac.id

Abstrak

Covid-19 telah ditetapkan sebagai Pandemi oleh World Health Organization (WHO). Salah satu antisipasi yang bisa dilakukan adalah dengan melakukan lockdown. Pada penelitian ini, akan disampaikan mengenai pembuatan pemodelan prediksi terkait analisa sentimen terkait “Lockdown” pada media sosial Twitter. Metode yang digunakan adalah dengan melakukan labeling menggunakan Vader dan selanjutnya tweet dilakukan ekstraksi menggunakan TF-IDF, dan dibuatkan pemodelan untuk prediksi sentimen menggunakan Naïve Bayes dan Support Vector Machine. Hasilnya yang didapat dari kedua algoritma tersebut ialah lebih dari 80%.

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

Abstract

Covid-19 has been set as a Pandemic by the World Health Organization (WHO). One anticipation that can be done is to do lockdown. In this study, will be presented the predictive modeling for sentiment analysis related to "lockdown" 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

1. PENDAHULUAN

Novel coronavirus 2019 (nCoV-2019), atau yang sekarang disebut dengan coronavirus disease-19 (COVID-19) sudah ditetapkan sebagai Pandemi oleh World Health Organization (WHO) [1].

Hal ini berakibat banyak negara mulai melakukan “lockdown” untuk mengantisipasi penyebaran virus ini [2]. Di Indonesia, hal ini belum dilakukan oleh Pemerintah walaupun terdapat beberapa penduduk Indonesia menyuatkan “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 [3]. 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 menyuatkan untuk melakukan lockdown melalui Twitter. Hal inilah yang mendorong

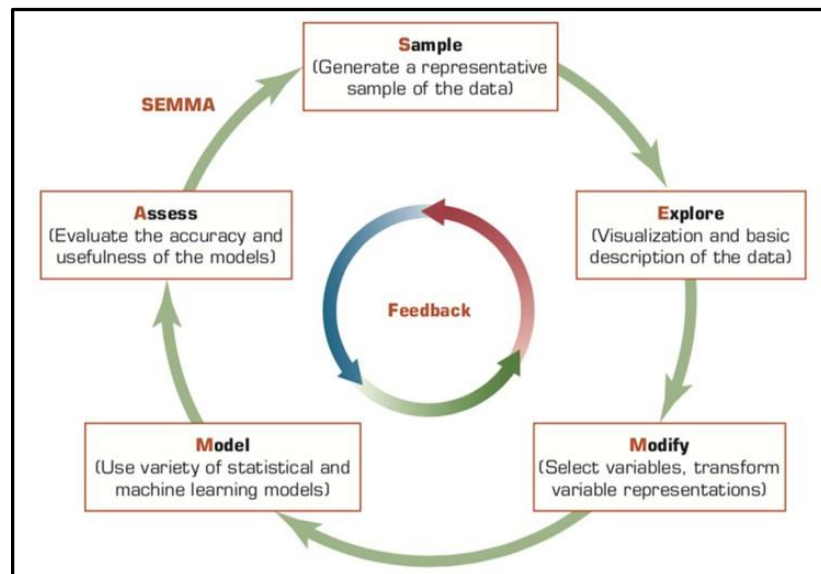
penelitian ini untuk membuat sebuah pemodelan analisis sentimen terkait dengan tweets mengenai “lockdown” di Indonesia.

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 mengelompokkan atau mengklasifikasi opini, penilaian seseorang terkait kreasi, organisasi, atau aktivitas tertentu [4]. Tujuan dari melakukan analisa sentimen ialah untuk menentukan perilaku dan opini penulis terhadap topik tertentu [5].

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 [6], [7], [8]. Pada penelitian kali ini, SVM dan Naïve-Bayes digunakan sebagai metode klasifikasi untuk memprediksi label sentimen pada suatu tweets pada twitter.

2. 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 [9]. Tahapan yang dilakukan pada penelitian ini dijelaskan pada Gambar 1.



Gambar 1. Semma Data Mining Process [10]

2.1 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.

2.2 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.

2.3 Modify

Pada tahap ini dilakukan beberapa metode persiapan data diantaranya:

2.3.1. Tokenisasi

Tokenisasi dilakukan untuk memecah sekumpulan kata (kalimat) menjadi kata yang memiliki arti tertentu.

2.3.2. 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

2.3.3. 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.

2.3.4. Stopword Removal

Tahap ini merupakan proses untuk melakukan filter terhadap kata – kata umum seperti “the”, “a”, dan lainnya, yang tidak diperlukan saat pemrosesan data.

2.3.5. 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.

2.4 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. Gambaran proses pada tahap Model digambarkan pada Gambar 2.

2.5 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.

3. 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.

3.1 Hasil pada tahap Sample & Explore

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.

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
4	I was supposed to have my first cum bukkake ne...
5	RT @KChiruTweets: On a Sunday before
6	RT @tanwer_m: This is from Jehanabad, Bihar. \...

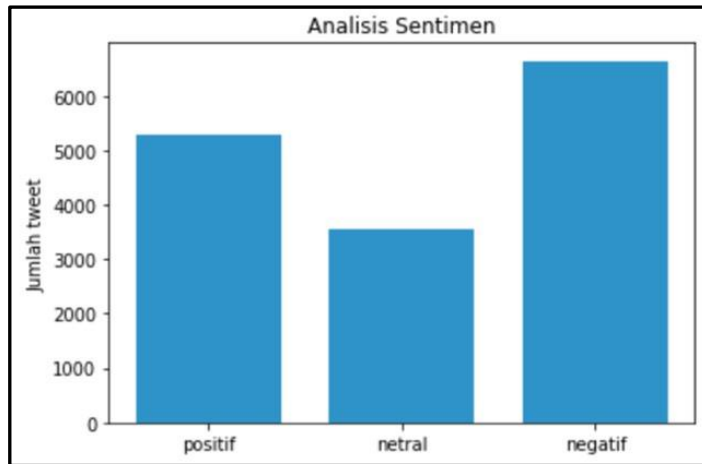
3.2 Hasil pada Modify

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

3.2 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).

3.2 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:

- Naïve Bayes

Pada Tabel 3 dan tabel 4 berturut-turut adalah hasil 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

Tabel 4. Classification Report (Naive Bayes)

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

- 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)

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

4. KESIMPULAN

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.

5. SARAN

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.

DAFTAR PUSTAKA

- [1] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomedica*. 2020.
- [2] H. Lau et al., "The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China," *J. Travel Med.*, 2020.
- [3] "Text Mining Dan Sentimen Analisis Twitter Pada Gerakan Lgbt," *Intuisi J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017.
- [4] G. A. Buntoro et al., "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 1–10, 2019.
- [5] F. Pramono, Didi Rosiyadi, and Windu Gata, "Integrasi N-gram, Information Gain, Particle Swarm Optimation di Naïve Bayes untuk Optimasi Sentimen Google Classroom," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 383–388, 2019.
- [6] Y. Ma and G. Guo, *Support vector machines applications*. 2014. [7] Bustami, "Penerapan Algoritma Naive Bayes," *J. Inform.*, 2014.
- [8] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, "Naive Bayes for Regression," *Mach. Learn.*, vol. 41, no. 1, pp. 5–25, 2000.

- [9] A. Azevedo and M. F. Santos, “KDD, semma and CRISP-DM: A parallel overview,” in MCCSIS’08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008, 2008.
- [10] P. A. Wardhani, Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th ed., vol. 6. Pearson, 2015