

Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial

Ridho Prabowo^{a1}, Herry Sujaini^{b2}, Tedy Rismawan^{a,b3}

^aJurusan Rekayasa Sistem Komputer Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, Kalimantan Barat 78115

¹ridhoprbwo@student.untan.ac.id

³tedyrismawan@siskom.untan.ac.id

^bJurusan Informatika Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, Kalimantan Barat 78115

²hs@untan.ac.id

Abstrak

Seiring berkembangnya *Coronavirus Disease 2019* (COVID-19) di Wuhan dan berkembang di seluruh negara khususnya di Indonesia. Banyak masyarakat di Indonesia menuangkan dan mengekspresikan opini mereka pada media sosial salah satunya Twitter. Opini tersebut menjadi acuan pada penelitian ini untuk menganalisis sentimen masyarakat Indonesia terhadap kasus COVID-19. Analisis sentimen merupakan proses pengolahan data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung pada *tweet*. Implementasi sistem dilakukan dengan mengklasifikasi *tweet* menggunakan metode Regresi Logistik Multinomial. Metode Regresi Logistik Multinomial dikenal dengan regresi logistik dengan variabel dependen mempunyai skala nominal lebih dari dua kategori. *Tweet* diklasifikasi menjadi 3 kategori sentimen yaitu sentimen positif, netral dan negatif. Sistem analisis sentimen berbasis *website* menggunakan bahasa pemrograman *Python*. Proses analisis sentimen melalui beberapa tahapan, pertama proses *crawling* data *tweet*, kemudian hasil *crawling* akan diproses ke tahap *text preprocessing*, setelah melewati *text preprocessing* *tweet* akan dibobot menggunakan TF-IDF, kemudian *tweet* diklasifikasi dengan menggunakan metode Regresi Logistik Multinomial ke dalam kelas sentimen positif, netral, atau negatif. Pengujian dilakukan menggunakan 870 data *tweet* yang telah dilabel terlebih dahulu. Sistem analisis sentimen menghasilkan akurasi sebesar 64%, dengan *precision* untuk sentimen positif 85%, netral 56%, dan negatif, 53% dan *recall* untuk sentimen positif 74%, netral 67%, dan negatif 50%.

Kata kunci: covid-19, analisis sentimen, regresi logistik multinomial, *twitter*.

The Sentiment Analysis of Twitter User about COVID-19 Cases in Indonesia Using the Multinomial Logistics Regression Method

Abstract

As the Coronavirus disease 2019 (COVID-19) develops in Wuhan and develops throughout the countries, especially in Indonesia. Many people in Indonesia are pouring and expressing their opinions on social media, one of which is Twitter. Sentiment analysis is an automatic data processing process to obtain sentiment information contained in tweets. The implementation of the system is done by classifying tweets using the Multinomial Logistics Regression method. The Multinomial Logistics Regression method known as logistic regression with the dependent variable has a nominal scale of more than two categories. Tweets are classified into 3 sentiment categories, namely positive, neutral and negative sentiment. The sentiment analysis system is website-based uses the Python programming language. The sentiment analysis process goes through several stages, first the tweet data through crawling process, then the crawling results will through the text pre-processing stage, after passing the tweet text preprocessing it will be weighted using TF-IDF, then classified tweets using the Multinomial Logistics Regression method into positive, neutral, or negative sentiment class. The test was carried out using 870 tweet data that had been previously labeled. The sentiment analysis system produces an accuracy of 64%, with precision for positive sentiments of 85%, neutral sentiments of 56%, and negative sentiments of 53% and recall for positive sentiments of 74%, neutral sentiments of 67%, and negative sentiments of 50%.

Keywords: covid-19, sentiment analysis, multinomial logistic regression, *twitter*.

I. PENDAHULUAN

Coronavirus disease 2019 (COVID-19) merupakan penyakit infeksi saluran pernapasan yang disebabkan oleh *severe acute respiratory syndrome virus corona 2 (SARS-CoV-2)*, atau sering disebut virus Corona [1]. Virus penyebab COVID-19, Sars-CoV-2 terus mengalami mutasi yang membentuk varian baru. Yang paling baru dan sudah terdeteksi di banyak negara adalah varian Omicron. Varian ini pertama kali dilaporkan di Afrika Selatan pada tanggal 24 November 2021 dan saat ini telah menyebar ke seluruh dunia [2]. Berkembangnya varian Omicron memicu kalangan masyarakat memposting ataupun menuangkan pendapat mereka secara masif terkait COVID-19 khususnya pada sosial media *twitter*. *Twitter* merupakan sebuah situs web yang dimiliki dan dioperasikan oleh *twitter inc.* Pesatnya perkembangan media sosial diberbagai bidang membuat media sosial tidak hanya sebagai wadah untuk bersosialisasi dan berkomunikasi tetapi juga sebagai wadah untuk menyampaikan opini tentang apa saja yang terjadi dan dirasakan oleh masyarakat. Penyampaian opini tersebut menjadi acuan untuk mengetahui sentimen masyarakat Indonesia terhadap kasus COVID-19.

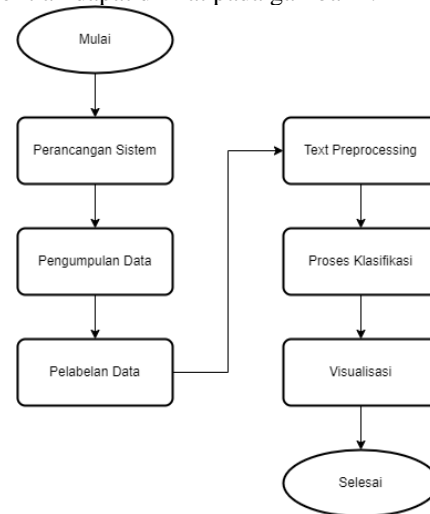
Analisis sentimen adalah proses penggunaan analitik teks untuk mendapatkan berbagai sumber data dari internet dan beragam *platform* media sosial. Tujuannya untuk memperoleh opini dari pengguna yang terdapat pada *platform* tersebut [3]. Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing (NLP)* yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Analisis sentiment dapat dikategorikan dalam aktivitas klasifikasi, yaitu mengelompokkan sentimen dalam sebuah data teks ke dalam kelasnya [4]. Untuk menentukan nilai kelas sentimen digunakan algoritma regresi logistik multinomial. Pemilihan algoritma ini dikarenakan algoritma regresi logistik multinomial merupakan algoritma yang digunakan ketika variabel dependen mempunyai skala lebih dari dua kategori.

Penelitian ini mencoba untuk mengklasifikasikan sentimen pengguna *twitter* terkait kasus COVID-19 di Indonesia. Adapun tahapan-tahapan untuk melakukan analisis sentimen tersebut, yang pertama adalah *crawling* data untuk menarik *tweet* pengguna *twitter* yang membahas tentang COVID-19. Setelah didapatkan *tweet* melalui *crawling* dilakukan *text preprocessing* untuk mengurangi *noise* yang terdapat *tweet*. *Text preprocessing* terdiri dari *casefolding*, tokenisasi, *filtering* dan *stemming*. Kemudian data yang telah dilakukan *text preprocessing* akan dilakukan pembobotan menggunakan *Term Frequency - Inverse Document Frequency (TF-IDF)* terlebih dahulu. Setelah didapatkan hasil pembobotan data *tweet* akan diklasifikasikan menggunakan metode regresi logistik multinomial untuk menentukan kelas sentimen apakah *tweet* tersebut bersifat positif, netral atau negatif.

II. METODOLOGI

Penelitian ini menganalisis sentimen pengguna *twitter* terhadap kasus COVID-19 di Indonesia dengan mengaplikasikan metode regresi logistik multinomial.

Metodologi berisi seluruh tahapan penelitian mulai dari metode pengumpulan data, proses klasifikasi sampai dengan visualisasi hasil analisis sentimen. Semua tahapan alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Diagram alir penelitian

A. Perancangan Sistem

Perancangan sistem dilakukan pada penelitian ini agar mendapatkan hasil sesuai dengan studi kasus, perancangan sistem dimulai dengan membuat alur sistem, basis data dan rancangan antarmuka aplikasi agar mempermudah interaksi terhadap sistem.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data *tweet* pengguna *twitter* yang membahas tentang COVID-19 di Indonesia. Data yang digunakan adalah data tekstual yang di *crawling* dari media sosial *twitter*. Data yang sudah didapatkan dari hasil *crawling* akan menjadi *dataset* yang akan digunakan untuk melatih dan menguji data. Data latih masing-masing dibagi dan diberi label sentimen positif, netral, dan negatif terlebih dahulu, kemudian akan digunakan untuk melatih sistem. Dari data latih tersebut digunakan juga untuk evaluasi sistem menggunakan *threshold 10-cross validation*. *Dataset* yang digunakan untuk data latih sebanyak 870 data *tweet* yang telah dilabel dengan masing-masing kelas positif, netral, dan negatif sebanyak 290 data *tweet*.

C. Pelabelan Data

Pelabelan data adalah proses melabel data ke dalam kelas-kelas yang telah ditentukan. Pelabelan dilakukan secara manual dengan membaca setiap ulasan *tweet* pengguna *twitter*, kemudian ditentukan ulasan tersebut dikategorikan kelas sentimen positif, netral atau negatif.

D. Text Preprocessing

Text preprocessing merupakan tahapan awal dari *text mining*. Tahap ini mencakup semua rutinitas dan proses untuk mempersiapkan data yang akan digunakan pada proses *knowledge discovery* sistem *text mining* [5]. Proses pembersihan data *tweet* dengan *text preprocessing*

bermaksud untuk mengurangi *noise* pada data *tweet* tersebut. Tahapan *text preprocessing* meliputi *case folding*, tokenisasi, *filtering*, dan *stemming*. Dengan dilakukannya *text preprocessing* akan terbentuk *dataset* bersih, *dataset* yang terbentuk dari proses ini akan memudahkan dalam pemrosesan sistem [6].

1) *Casefolding*: Tahapan pertama dalam *text preprocessing* yaitu *casefolding*. *Casefolding* merupakan proses mengubah atau mengkonversi setiap karakter didalam kata atau kalimat menjadi huruf kecil [7].

2) *Tokenisasi*: Tahapan yang dilakukan setelah *casefolding* adalah tokenisasi. Tokenisasi merupakan proses untuk memecah kata, serta melakukan penghapusan delimeter seperti titik (.), koma (,), spasi, dan karakter angka pada kalimat [8].

3) *Filtering*: Setelah dilakukan tokenisasi tahapan selanjutnya yaitu *filtering*. *Filtering* merupakan proses yang dilakukan untuk untuk menghapus kata yang tidak memiliki arti/makna dari hasil tokenisasi. Proses dalam *filtering* dalam membuang kata-kata yang tidak digunakan atau *stopword* terdapat dalam *stoplist* [9].

4) *Stemming*: Tahapan terakhir dari *text preprocessing* pada penelitian ini adalah *stemming*. *Stemming* merupakan proses menghilangkan imbuhan-imbuhan baik itu prefix, sufiks, maupun konfiks yang pada setiap kata [10].

E. Proses Klasifikasi

Proses klasifikasi merupakan proses untuk mengelompokkan data *tweet* ke dalam kelas-kelas sentimen positif, netral, dan negatif. Di dalam proses klasifikasi ini menggunakan perhitungan TF-IDF untuk pembobotan kata, dan metode regresi logistik multinomial untuk klasifikasi *tweet*.

1) Pembobotan kata TF-IDF: Metode TF-IDF merupakan metode untuk menghitung setiap kata yang paling umum digunakan pada *information retrieval* [11]. Pemberian bobot kata atau *term* pada setiap dokumen mencerminkan betapa pentingnya sebuah *term* pada dokumen [12]. Dokumen yang dimaksud pada penelitian ini adalah *tweet*. Berikut adalah persamaan TF-IDF.

$$w_{ij} = t_{f_{ij}} \times idf_j \quad (1)$$

2) Regresi Logistik Multinomial: Analisis regresi logistik multinomial merupakan regresi logistik yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat atau mempunyai skala yang bersifat polychotomous atau multinomial yaitu berskala nominal dengan lebih dari dua kategori [13],[14]. Bentuk umum fungsi logit dengan variabel respon terdiri dari tiga kategori sebagai berikut.

$$g_j(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p \quad (2)$$

Suatu variabel respon dengan tiga kategori akan membentuk dua persamaan logit, di mana masing masing persamaan ini membentuk regresi logistik multinomial yang membandingkan suatu kelompok kategori terhadap pembanding, yaitu sebagai berikut.

$$g_1(x) = \log \frac{P(Y=2|x)}{P(Y=1|x)} = \log \frac{\pi_2(x)}{\pi_1(x)} = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p \quad (3)$$

$$g_2(x) = \log \frac{P(Y=3|x)}{P(Y=1|x)} = \log \frac{\pi_3(x)}{\pi_1(x)} = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p \quad (4)$$

Berdasarkan kedua peluang kumulatif pada persamaan (3) dan (4), didapatkan peluang untuk masing-masing kategori respon sebagai berikut.

$$P(Y=1|x) = \pi_1(x) = \frac{\exp g_1(x)}{1 + \exp g_1(x) + \exp g_2(x)} \quad (5)$$

$$P(Y=2|x) = \pi_2(x) = \frac{\exp g_2(x)}{1 + \exp g_1(x) + \exp g_2(x)} \quad (6)$$

$$P(Y=3|x) = \pi_3(x) = \frac{1}{1 + \exp g_1(x) + \exp g_2(x)} \quad (7)$$

3) *Confusion Matrix*: *Confusion Matrix* digunakan untuk evaluasi model yang digunakan pada *machine learning* [15]. Setiap unsur matriks menunjukkan jumlah contoh data uji untuk kelas sebenarnya yang digambarkan dalam bentuk baris sedangkan kolom menggambarkan kelas yang di prediksi. Tabel confusion matrix dapat dilihat pada Tabel 1.

TABEL 1
CONFUSION MATRIX

Nilai Aktual	Nilai Prediksi		
	Positif	Negatif	Netral
Positif	TPos	FPosNeg	FPosNet
Negatif	FNegPos	TNeg	FNegNet
Netral	FNetPos	FNetNeg	TNet

Dari Tabel 1 didapatkan persamaan untuk akurasi, *precision*, dan *recall*.

a. Akurasi

$$\bullet AC = \frac{TPos + TNeg + TNet}{Total Data} \times 100\% \quad (8)$$

b. *Precision*

$$\bullet Pre(Pos) = \frac{TPos}{TPos + FNegPos + FNetPos} \times 100\% \quad (9)$$

$$\bullet Pre(Neg) = \frac{TNeg}{TNeg + FNegPos + FNetNeg} \times 100\% \quad (10)$$

$$\bullet Pre(Net) = \frac{TNet}{TNet + FPosNet + FNetNet} \times 100\% \quad (11)$$

c. *Recall*

$$\bullet Rec(Pos) = \frac{TPos}{TPos + FPosNeg + FPosNet} \times 100\% \quad (12)$$

$$\bullet Rec(Neg) = \frac{TNeg}{TNeg + FNegPos + FNegNet} \times 100\% \quad (13)$$

$$\bullet Rec(Net) = \frac{TNet}{TNet + FNetPos + FNetNeg} \times 100\% \quad (14)$$

F. Visualisasi

Hasil dari proses klasifikasi akan di visualisasi dalam bentuk aplikasi berbasis *website* yang menampilkan *section* halaman awal, dan *section* halaman hasil yang berisi *barchat*, dan *tweet* beserta sentimennya dalam bentuk table selama rentang waktu 7 hari terakhir.

III. HASIL DAN PEMBAHASAN

A. Hasil Text Preprocessing

Sebelum data *tweet* diklasifikasi, ada tahapan yang harus dilalui terlebih dahulu, yaitu *text preprocessing*. Setelah data *tweet* diperoleh dari hasil *crawling*, maka data *tweet* akan dilakukan *text preprocessing*. Contoh data *tweet* pada Gambar 2.

Sebanyak 186 pelajar SMA Matauli, Tapanuli Tengah, Sumatra Utara, terpapar Covid-19 varian Omicron. Kegiatan PTM dihentikan sementara. HEL

#SeputariNewsRCTI #Siswa #Tapanuli #COVID19
#TanpaGejala #Gejala #SumatraUtara
<https://t.co/p5YZ4Oi8CI>

Gambar 2. Contoh *tweet*

Tahapan *text preprocessing* meliputi *case folding*, tokenisasi, *filtering*, dan *stemming*.

1) *Case Folding*: *Case folding* dilakukan untuk menyeragamkan semua karakter pada *tweet* menjadi huruf kecil. Contoh hasil *casefolding* data *tweet* pada Gambar 3.

sebanyak 186 pelajar sma matauli, tapanuli tengah, sumatra utara, terpapar covid-19 varian omicron. kegiatan ptm dihentikan sementara. hel

#seputarinewsrtci #siswa #tapanuli #covid19
#tanpagejala #gejala #sumatrautara
<https://t.co/p5yz4oi8ci>

Gambar 3. Contoh hasil *case folding*

2) Tokenisasi: Proses tokenisasi memecah kata-kata dari *tweet* dan penghapusan delimiter *tweet entity*, *hastag*, *retweet* dan *mention*. Contoh hasil tokenisasi data *tweet* pada Tabel 2.

TABEL II
CONTOH HASIL PROSES TOKENISASI DATA TWEET

sebanyak	pelajar	sma	matauli	tapanuli	tengah
sumatra	utara	terpapar	covid	varian	omicron
kegiatan	ptm	dihentikan	sementara	hel	

3) *Filtering*: Proses *filtering* mengambil kata-kata penting dari hasil tokenisasi. Proses *filtering* dilakukan dengan menghapus kata-kata yang tidak memiliki makna yang terdaftar di *stopwords*. Contoh hasil *filtering* data *tweet* pada Tabel 3.

TABEL III
CONTOH HASIL PROSES FILTERING DATA TWEET

pelajar	sma	matauli	tapanuli	sumatra	utara	terpapar
covid	varian	omicron	kegiatan	ptm	dihentikan	hel

4) *Stemming*: Proses *stemming* mengembalikan kata-kata dari hasil *filtering* ke kata-kata bentuk dasar. Contoh hasil *stemming* data *tweet* pada Tabel 4.

TABEL IV
CONTOH HASIL PROSES STEMMING DATA TWEET

ajar	sma	matauli	tapanuli	sumatra	utara	papar
covid	varian	omicron	giat	ptm	henti	hel

Setelah semua tahapan *text preprocessing* selesai, data *tweet* sudah siap untuk diproses ke tahapan klasifikasi.

B. Pengujian

Pengujian dilakukan untuk memastikan sistem yang dibuat sudah berfungsi sesuai dengan harapan, seperti penerapan algoritma kedalam sistem. Data yang digunakan berjumlah 870 data yang berisikan opini dari pengguna *twitter* terhadap kasus COVID-19 yang telah dilabel. Masing-masing 290 opini positif, 290 opini netral, dan 290 opini negatif. Sebagai contoh hasil pengujian algoritma regresi logistik multinomial menggunakan 10 data *tweet*. *Tweet* tersebut telah melalui proses *text preprocessing* dan pembobotan kata menggunakan TF-IDF. Hasil pengujian algoritma regresi logistik multinomial menggunakan bantuan *library scikit-learn* dengan *threshold 10-cross validation*. Tabel hasil klasifikasi dapat dilihat pada Tabel 5.

TABEL V
TABEL HASIL KLASIFIKASI

Tweet	Prob Negatif	Prob Netral	Prob Positif	Kelas
1	0.24790472	0.19561385	0.55648143	Positif
2	0.2535374	0.55615191	0.19031069	Netral
3	0.26976368	0.18940657	0.54082975	Positif
4	0.61567572	0.19855417	0.18577011	Negatif
5	0.26101674	0.54379165	0.19519161	Netral
6	0.24790472	0.19561385	0.55648143	Positif
7	0.62661057	0.1851426	0.18824683	Negatif
8	0.6271796	0.18785714	0.18496326	Negatif
9	0.60676945	0.18684898	0.20638156	Negatif
10	0.24363738	0.56101929	0.19534333	Netral

Setelah mendapatkan hasil klasifikasi dilakukan pengujian hasil klasifikasi dengan data *tweet* yang telah dilabel. Digunakan *confusion matrix* untuk mengukur akurasi, *precision*, dan *recall*. Hasil *confusion matrix* pengujian pada Tabel 6.

TABEL VI
CONFUSION MATRIX PENGUJIAN

Nilai Aktual	Nilai Prediksi		
	Positif	Negatif	Netral
Positif	216	32	42
Negatif	11	193	86
Netral	28	117	145

Akurasi yang dimaksud adalah persentase tingkat kedekatan antara nilai yang didapat terhadap nilai sebenarnya. Akurasi menggunakan metode regresi logistik diperoleh sebesar 64%. Kemudian didapatkan *precision* dan *recall* untuk masing-masing sentimen. *Precision* yang dimaksud adalah persentase tingkat ketepatan antara bagian data yang diambil dengan informasi yang dibutuhkan. Untuk *Precision* sentimen positif memperoleh 85%, *precision* sentimen netral memperoleh 56%, dan *precision* sentimen negatif memperoleh 53%. *Recall* yang dimaksud adalah persentase tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Recall* sentimen positif memperoleh 74%, *recall* sentimen netral memperoleh 56%, dan *recall* sentimen negatif memperoleh sebesar 50%. Hasil pengujian akurasi, *precision*, dan *recall* pada Tabel 7.

TABEL VII
HASIL AKURASI, *PRECISION*, DAN *RECALL*

Sentimen	Akurasi	<i>Precision</i>	<i>Recall</i>
Positif	64%	85%	74%
Netral		56%	67%
Negatif		53%	50%

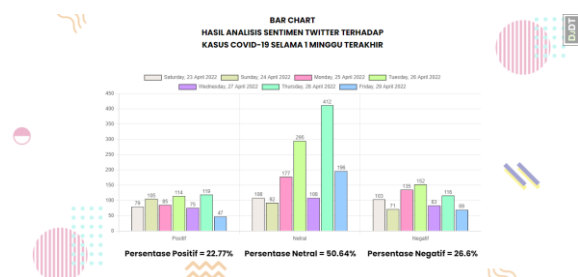
C. Implementasi Sistem

Implementasi sistem dilakukan menerapkan metode yang digunakan kedalam sistem. Sistem yang dibuat berbasis aplikasi *website* dengan menerapkan bahasa pemrograman Python dan *framework* Django.



Gambar 4. Section Landing Page

Gambar 4 merupakan halaman pertama dari sistem ini. Halaman ini menampilkan animasi analisis sentimen serta judul dari aplikasi. Pada Halaman ini terdapat tombol yang mengarah ke *section* hasil analisis sentimen untuk melihat hasil analisis.



Gambar 5. Bar Chart Hasil Analisis Sentimen

Gambar 5 merupakan hasil dari analisis sentimen berupa *bar chart* yang menampilkan jumlah sentimen setiap harinya selama 1 minggu terakhir beserta persentase masing-masing sentimen.

TABEL
HASIL ANALISIS SENTIMEN TWITTER TERHADAP KASUS COVID-19 SELAMA 1 MINGGU TERAKHIR

Showing 1 to 25 of 2746 entries

No.	Tweet	Date	Sentiment
1	RT @BISU... pating Komite Penanganan Corona Virus Disease 2019 dan Pemulihan Ekonomi Nasional (KPCPEN) melansir data, hingga Rabu (27/4), tot...	Apr 29	Netral
2	Komite Penanganan Corona Virus Disease 2019 dan Pemulihan Ekonomi Nasional (KPCPEN) melansir data, hingga Rabu (27/4), total cakupan vaksinasi Indonesia mencapai 405,9 juta suntikan, dari target 400 juta. https://t.co/9tHvRdK...	Apr 29	Positif
3	RT @g88CIndonesia: Lebih dari 80 juta orang dipekerjakan pulang kampung pada Sabtu (tupun ini, setelah pemerintah untuk pertama kalinya)...	Apr 29	Negatif
4	RT @g88CIndonesia: Update Infografi persentase penanganan COVID-19 di Indonesia per tanggal 29 April 2022 Pukul 12.00 WIB. @beritasawan...	Apr 29	Netral
5	RT @g88CIndonesia: Update Infografi persentase penanganan COVID-19 di...	Apr 29	Netral

Gambar 6. Tabel Hasil Analisis Sentimen

Gambar 6 merupakan hasil dari analisis sentimen berbentuk tabel yang menampilkan *tweet* beserta tanggal dan sentimen dari *tweet* tersebut dalam bentuk tabel. Pada halaman ini bisa memilih hari pada rentang waktu selama 1 minggu terakhir.

IV. KESIMPULAN

Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa sistem dapat mengklasifikasi *tweet* terhadap kasus COVID-19 di Indonesia secara otomatis ke dalam sentimen positif, netral, dan negatif. Adapun data latih yang digunakan sebanyak 870 yang telah dilabel dengan masing-masing data sentimen positif 290, sentimen netral 290, dan sentimen negatif 290 yang diuji menggunakan *threshold 10-cross validation*. Sistem ini memiliki akurasi sebesar 64% dengan *precision* positif sebesar 85%, netral sebesar 56%, dan negatif sebesar 53% serta *recall* positif sebesar 74%, netral sebesar 67%, dan negatif sebesar 50%.

Adapun faktor yang mempengaruhi akurasi model regresi logistik yaitu pelabelan dan *text preprocessing*. Dimana pelabelan dilakukan secara manual dan data *slang word* pada *text preprocessing* harus selalu ditambah mengingat banyaknya kosakata yang tidak baku yang digunakan pengguna *twitter* saat membuat *tweet*.

DAFTAR PUSTAKA

- [1] Yuliana, "Corona Virus Diseases (Covid-19): Suatu tinjauan literatur.," *Wellness And Healthy Magazine*, no. Universitas Aisyah Pringsewu, p. 187–192, 2020.
- [2] Kominfo RI, "Tujuh hal yang perlu diketahui dari varian Omicron penyebab COVID-19," Kominfo RI, 29 November 2021. [Online]. Available: <https://covid19.go.id/edukasi/masyarakat-umum/7-hal-yang-perlu-diketahui-dari-varian-omicron-penyebab-covid-19>.
- [3] A. Syahadati, "Analisis Sentimen Penerapan PSBB Di DKI Jakarta dan Dampaknya Terhadap Pergerakan IHSG," *Jurnal TEKNOINFO*, vol. 15, No. 1, pp. 20–25, 2021.
- [4] A. N. M. P. B. S. G. 2. Basant, "Sentiment Analysis Using Common-Sense and Context Information," no. Hindawi Publishing Corporation Computational Intelligence and Neuroscience, 2015.

- [5] R. Feldman dan J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, New York: Cambridge University Press, 2007.
- [6] L. Ardiani, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)* Vol. 8 , No. 2 , April 2020, pp. 183-190, 2020.
- [7] S. Adi, "Perancangan Klasifikasi Tweet Berdasarkan Sentimen dan Fitur Calon Gubernur DKI Jakarta 2017," *Journal Of Informatic Pelita Nusantara*, Vol. 3 No. 1, 2018.
- [8] D. Pakpahan dan H. Widyastuti, "Aplikasi Opinion Mining dengan Algoritma Naïve Bayes untuk Menilai Berita Online," *Jurnal Integrasi*, vol. 6, no. 1, pp. 1-10, 2014.
- [9] E. E. Pratama and B. R. Trilaksono, "Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)," *J. Edukasi dan Penelit. Inform*, 2015.
- [10] F. Z. Tala, *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*, Amsterdam: Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands, 2003.
- [11] S. Robertson, "Understanding Inverse Document Frequency: On theoretical arguments for IDF," *Journal of Documentation* 60 no. 5, p. 503–520.
- [12] M. Yunus, "Medium," 30 April 2020. [Online]. Available: <https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>.
- [13] Hosmer dan Lemeshow, *Applied Logistic Regression*. Second Edition, New York: John Wiley & Sons, 2000.
- [14] A. Agresti, *Categorical Data Analysis*, New York: John Wiley and Sons, Inc., 1990.
- [15] C. D. R. P. & S. H. Manning, *An Introduction to Information Retrieval – Online Edition*, Cambridge: Cambridge University Press, 2009.