Analisis Sentimen Twitter: Penanganan Pandemi Covid-19 Menggunakan Metode *Hybrid Naïve Bayes, Decision Tree, dan Support Vector Machine*

1st Muyassar Akmal Iftikar
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
akmaliftikar@student.telkomuniversity.ac.id

2nd Yuliant Sibaroni
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
yuliant@telkomuniversity.ac.id

Abstrak

Indonesia menduduki peringkat keenam sebagai negara dengan pengguna media sosial Twitter terbanyak.Pada bulan Desember 2019, sebuah virus yang dijuluki COVID-19 muncul dan mulai menyebar ke hampirseluruh penjuru dunia. Penulis bermaksud untuk melakukan analisis sentimen pengguna Twitter tentang penanganan COVID-19 di Indonesia serta membandingkan hasil akurasi dari model hybrid (*Stacking Ensemble)* dan model-model individu lainnya. Model-model klasifikasi machine learning individu mempunyai kelebihan dan kekurangannya masing-masing, dan setiap modelnya mempunyai karakteristik yang berbeda dalam menjalankan proses klasifikasi. Penulis menggunakan *Stacking Ensemble* sebagai model klasifikasi hybrid. *Stacking Ensemble* bekerja dengan cara mengkombinasikan hasil prediksi dari model klasifikasi lainnya. Kemudian hasil tersebut akan dikombinasikan dengan meta-classifier (Logistic Regression) dengan tujuan untuk mendapatkan hasil prediksi akhir yang lebih akurat dibanding dengan hasil klasifikasi model tunggal. Dari penelitian ini, ditemukan bahwa reaksi umum pengguna Twitter di Indonesia terhadap penanganan COVID-19 di Indonesia secara umum adalah positif, dengan presentase sentimen positif 75.3% dan 60.39%. Sehingga, dapat disimpulkan bahwa penganganan COVID-19 di Indonesia dianggap baik oleh masyarakat Indonesia. Selain itu, ditemukan bahwa metode hybrid Stacking Ensemble dapat meningkatkan nilai akurasi yang dihasilkan oleh *classifiers* individu lainnya, dengan perbedaan 0.63% dan 1.02%.

Kata kunci: analisis sentimen, COVID-19, hybrid, Stacking Ensemble

Abstract

Indonesia is ranked sixth as the country with the most Twitter social media users. In December 2019, a viruscalled COVID-19 emerged and spread to almost all over world. The author intends to do sentiment analysisof Twitter users regarding the handling of COVID-19 in Indonesia and compare the accuracy results of the hybrid model (Stacking Ensemble) and other individual models. Machine learning classification models have advantages and disadvantages of their own, and each model has different characteristics in carrying out the classification process. The author uses the Stacking Ensemble as a hybrid classification model. Stacking Ensemble works by combining prediction results from other classification models. Then these results will be combined with a meta-classifier (Logistic Regression) with the aim of getting a final prediction result that is more accurate than the results of a single model classification. From this study, it was found that the general reaction of Twitter users in Indonesia to the handling of COVID-19 in Indonesia was generally positive, with a positive sentiment percentage of 75.3% and 60.39%, respectively. So, it can be concluded that Indonesian citizens' response to the handling of COVID-19 in Indonesia is positive. In addition, it was found that the hybrid Stacking Ensemble method can increase the accuracy values produced by other individual classifiers, with a difference of 0.63% and 1.02%.

Keywords: sentiment analysis, COVID-19, hybrid, Stacking Ensemble

I. PENDAHULUAN

A. Latar Belakang

Twitter merupakan jejaring sosial media yang memungkinkan pengguna mengekspresikan perasaan dan pendapatnya tentang topik tertentu. Menurut [1], Indonesia menduduki peringkat keenam sebagai negara dengan pengguna media sosial Twitter terbanyak. Hal ini menunjukkan bahwa pengguna aktif Twitter di Indonesia memiliki peran yang signifikan dalam hal *engangement*.

Pada bulan Desember 2019, sebuah virus yang dijuluki COVID-19 muncul dan mulai menyebar ke hampir seluruh penjuru dunia, tak terkecuali di Indonesia. Dilansir oleh WHO hingga bulan Februari 2021, tercatat bahwa Indonesia memiliki angka kasus mencapai 4.3 juta kasus positif, dan 144.320 kematian [2]. Dengan angka tersebut, dapat disimpulkan bahwa COVID-19 memiliki pengaruh besar terhadap kehidupan masyarakat Indonesia.

Tidak sedikit dari pengguna Twitter di Indonesia memanfaatkan media sosial Twitter untuk dijadikan sebagai wadah berbagi informasi serta sarana untuk mengemukakan pendapat yang berkaitan dengan persebaran,penanganan, serta upaya pemerintah Indonesia dalam menghadapi COVID-19. Reaksi dari penggunan Twitter pun beragam. Informasi dan reaksi tersebut, terlepas dari kebenaran dan keabsahannya, dapat dilihat melalui *tweet- tweet* dari berbagai akun, terutama akun-akun yang memiliki pengikut yang tinggi, seperti pejabat pemerintahan, lembaga kesehatan nasional/internasional, selebritas, dan lain sebagainya. Dengan pertimbangan tersebut, Twittermenjadi sarana yang kaya akan informasi,

yang dimanfaatkan peneliti-peneliti untuk melakukan analisis terhadappandemi ini. [3]

Melalui sosial media Twitter, penulis bermaksud untuk melakukan analisis sentimen pengguna Twitter tentang penanganan COVID-19 di Indonesia. Analisis Sentimen memanfaatkan salah satu proses dari *Natural Language Processing* untuk menilai emosi dan rasa seseorang, dalam hal ini pengguna Twitter, serta menentukan *Tweet* yang diposting masuk ke dalam kategori positif atau negatif [4]. Untuk mengklasifikasikan sebuah *tweet* masuk ke dalam kategori postitif atau negatif, dapat dilakukan dengan menggunakan model klasifikasi *machine learning.* Penulis menggunakan model Multinomial Naïve Bayes, *Decision Tree*, dan *Support Vector Machine.* Model-model klasifikasi *machine learning* tersebut mempunyai kelebihan dan kekurangannya masing-masing, dan setiap modelnya mempunyai karakteristik yang berbeda dalam menjalankan proses klasifikasi [5]. Oleh karena itu, satu model *machine learning* saja tidak dapat dijadikan model yang definitif dalam menyelesaikan proses klasifikasi.

Dari pertimbangan tersebut, penulis menggunakan *Stacking Ensemble* sebagai model klasifikasi *hybrid*. *Stacking Ensemble* bekerja dengan cara mengkombinasikan hasil prediksi dari masing-masing *base-classifiers* (Multinomial Naïve Bayes, *Decision Tree*, dan *Support Vector Machine).* Kemudian hasil tersebut akan dikombinasikan dengan *meta-classifier* (*Logistic Regression*) dengan tujuan untuk mendapatkan hasil prediksi akhir yang lebih akurat dibanding dengan hasil klasifikasi model tunggal [6]. Pada penelitian ini, penulis bermaksud untuk menentukan reaksi umum pengguna Twitter terhadap penanganan COVID-19 di Indonesia serta membandingkan hasil akurasi dari model hybrid (*Stacking Ensemble)* dan model-model individu lainnya (Multinomial Naïve Bayes, *Support Vector Machine, dan Decision Tree*).

B. Topik dan Batasannya

Topik yang dibahas dalam tugas akhir ini adalah bagaimana reaksi umum pengguna Twitter terhadap penanganan COVID-19 di Indonesia serta membandingkan hasil akurasi dari model hybrid (*Stacking Ensemble)* dan model-model individu lainnya (Multinomial Naïve Bayes, *Support Vector Machine, dan Decision Tree*).

Batasan masalah pada tugas akhir ini adalah sebagai berikut:

a. Dataset tweet sebanyak 3.000 *tweets* yang diterjemah dari Bahasa Indonesia ke Bahasa Inggris dan diberilabel secara otomatis dengan menggunakan model VADER dari *libra*ry NLTK
b. Dataset tweet sebanyak 3.000 *tweets* yang diberi label secara manual dan menggunakan *libra*ry Sastrawi
c. *Tweets* berkaitan dengan informasi dan opini penanganan COVID-19 di Indonesia
d. *Tweets* yang diunggah pada bulan Januari 2021 hingga bulan Desember 2021.

C. Tujuan

Tujuan dari tugas akhir ini adalah menentukan reaksi umum pengguna Twitter terhadap penanganan COVID-19 di Indonesia dan membandingkan hasil akurasi dari model hybrid (*Ensemble Stacking)* dan model-model individu lainnya (Multinomial Naïve Bayes, Support Vector Machine, dan Decision Tree).

D. Organisasi Tulisan

Struktur penulisan dari tugas akhir ini disusun sebagai berikut: Bagian pertama berisi pendahuluan menyangkut tugas akhir ini. Bagian kedua menjelaskan studi yang terkait dengan tugas akhir ini. Bagian ketiga menjabarkan pemodelan dari sistem yang dibangun dan data yang digunakan. Bagian keempat menjelaskan hasildan evaluasi hasil pengujian yang telah dilakukan pada bagian ketiga. Pada bagian terakhir menjelaskan kesimpulan dan saran berdasarkan hasil pengujian yang dilakukan pada tugas akhir ini.

II. KAJIAN TEORI

Pada banyak penelitian analisis sentimen, *Stacking Ensemble* merupakan model yang digunakan untuk menggabungkan metode-metode klasifikasi untuk menghasilkan hasil dengan akurasi yang maksimal. Sebagai contoh, pada [7] penelitinya menggunakan *Convolutional Neural Network* (CNN) sebagai *base-classifers* dan *Multi-Layer Perceptron (MLP)* sebagai *meta-classifiers.*

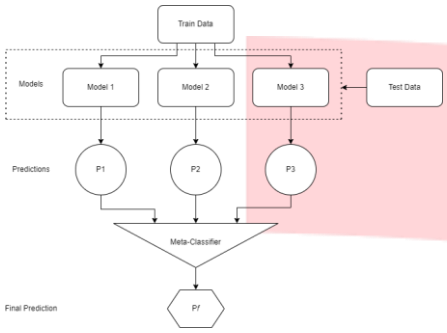
A. Setiment Analysis

Sentiment Analysis atau analisis sentimen merupakan penelitian yang menganalisis opini, sentimen, evaluasi, sikap, dan perasaan orang terhadap suatu topik, peristiwa, masalah, dan lain sebagainya [8]. Analisis sentimen digunakan untuk mengklasifikasi data berbasis teks ke dalam kelas. Kelas tersebut dibagi menjadi dua bagian, positif dan negatif. Jika data teks mengandung opini positif maka data tersebut memiliki polaritas positif dan dimasukkan ke dalam kelas positif. Sedangkan jika data teks mengandung opini negatif maka data tersebut memiliki polaritas negatif dan dimasukkan ke dalam kelas negatif. Pada analisis sentimen terdapat beberapa proses yaitu *pre-processing data, feature extraction,* klasifikasi, dan evaluasi hasil. *Pre-processing data* terdiri dari *data cleaning* dan *data preparation.* Selanjutnya, dilakukan *feature extraction* yang mengubah data teks menjadi data berbentuk vektor agar dapat diproses oleh model klasifikasi *machine learning* dan menghasilkan akurasi yang lebih akurat [9]. Di salah satu penelitian [10], peneliti melakukan analisis sentimen dengan menggunakan model Naïve Bayes, *Decision Tree*, dan *Random Forest,* dan hasilnya adalah bahwa metode Naïve Bayes menghasilkan akurasi paling tinggi dibanding dengan metode lainnya.

B. Stacking Ensemble

Algoritma *Stacking Ensemble* merupakan sebuah algoritma *machine learning* yang memungkinkan untuk menyelesaikan permasalahan dengan lebih dari satu model. *Stacking Ensemble* digunakan untuk mengkombinasikan metode-metode untuk membuat prediksi akhir [11]. Pada dasarnya, metode *ensemble learning* mengimplikasikan

bahwa keluaran dari beberapa *classifiers* akan menghasilkan prediksi yang lebih akurat. Berbeda dengan pendekatan-pendekatan konvensional yang mempelajari satu hipotesa dari data *training*, metode *ensemble* menggabungkan sejumlah hipotesa pada proses *learning* [7]. Meskipun tidak dapat dijamin bahwa metode *ensemble* akan menghasilkan hasil yang lebih baik, namun penggabungan model-model klasifikasi akan meminimalisir kesalahan dalam pemilihan model klasifikasi yang kurang baik [12]. Berikut ilustrasi dari *StackingEnsemble:*



Gambar 1. Ilustrasi *Stacking Ensemble*

C. Naïve Bayes

Naïve Bayes merupakan model pengelompokkan yang mengkalkulasi probabilitas pada masing-masing kelas berdasarkan pembagian kata-kata tertentu dalam sebuah bacaan [13]. Naïve Bayes classifier digunakan sebagai classifier probabilistik. Classifier ini menggunakan beberapa campuran model yang dapat membangun probabilitas dari komponen yang terdiri dari teori Bayes untuk dijalankan sebagai classifier probabilistik. [14] Metode ini dianggap memiliki tingkat akurasi yang tinggi pada kasus analisis sentimen dan telah diuji dengan beberapa algoritma lain [13] [10]**.** Pada [15], metode Naïve Bayes merupakan salah satu metode yang baik karenaalgoritmanya yang sederhana, dan dapat mengurangi kompleksitas dari informasi mengingat informasi yang dikumpulkan berjumlah besar. Selain itu, pada [16] dikatakan bahwa algoritma Naïve Bayes dan *Support Vector Machine* merupakan model yang sering digunakan untuk memecahkan masalah pada analisis sentimen. Salah satu model yang termasuk model Naïve Bayes adalah Multinomial Naïve Bayes. Multinomial Naïve Bayes digunakanuntuk menghitung berapa kali sebuah kata muncul dalam suatu dokumen teks [17]. Adapun rumus dari Multinomial Naïve Bayes adalah sebagai berikut:

Tweet ‘n’ dengan polaritas ‘p’ dapat dikalkulasikan dengan [18]:

$$P(p|n) \text{ a } P(p) \prod_{1 \le k \le nd} P(t_k|p) \quad \dots \text{(i)}$$

Dimana P(t_k|p) merupakan probabilitas kondisional apakah suatu kata t_k muncul dalam sebuah *tweet* atau tidakdengan polaritas p yang dapat dikalkulasikan sebagai berikut:

Pada persamaan di atas count $(t_k|p)$ adalah jumlah suatu kata t_k yang muncul di sebuah *tweet* yang mempunyai polaritas p dan count (t_p) adalah jumlah *token* yang muncul di sebuah *tweet* dengan polaritas p. Selain itu, 1 dan |V| ditambahkan untuk menghindari kesalahan kalkulasi pada saat suatu kata tidak muncul sama sekali di sebuah *tweet.*

$$P(t_k|p) = \frac{count\ (t_k|p)+1}{count\ (t_k|p)+|V|} \quad \text{(ii)}$$

Pada persamaan di atas, count $(t_k|p)$ adalah jumlah suatu kata t_k yang muncul di sebuah *tweet* yang mempunyai polaritas p dan count (t_p) adalah jumlah *token* yang muncul di sebuah *tweet* dengan polaritas p. Selain itu, 1 dan |V| ditambahkan untuk menghindari kesalahan kalkulasi pada saat suatu kata tidak muncul sama sekali di sebuah *tweet.*

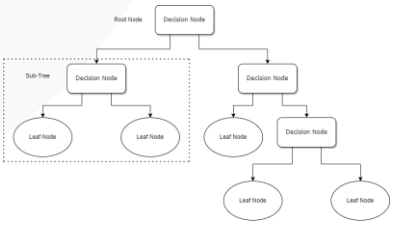
Pada persamaan di atas, count ($t_k|p$) adalah jumlah suatu kata t_k yang muncul di sebuah *tweet* yang mempunyai polaritas p dan count (t_p) adalah jumlah *token* yang muncul di sebuah *tweet* dengan polaritas p. Selain itu, 1 dan |V| ditambahkan untuk menghindari kesalahan kalkulasi pada saat suatu kata tidak muncul sama sekali di sebuah *tweet.*

P(p) merupakan probabilitas *tweet* sebelumnya dengan polaritas p yang dapat dikalkulasikan sebagai berikut:

$$P(p) = \frac{jumlah\ tweets\ dengan\ polaritas\ p}{total\ jumlah\ tweets} \quad \text{(iii)}$$

D. Decision Tree

Decision Tree merupakan model yang digambarkan sebagai sebuah pohon, yang terdiri dari node-node yang merepresentasikan atribut, untuk memprediksi hasil dari sebuah permasalahan. Node tertinggi dari sebuah *tree* disebut dengan *root node* [19]. Konstruksi dari *Decision Tree* dilakukan dengan cara mempartisi data rekursif. Pada setiap tahapnya, aturan terbaik *splitting* akan ditentukan, dan data dari *node* akan dibagi menjadi *child nodes* sesuai dengan kriteria tertentu. Prosedur yang sama dilakukan secara rekursif kepada seluruh *node* baru di dalam pohon yang dihasilkan sampai kondisi berhenti tercapai. Meskipkun dapat mempercepat dan memperjelas proses data *splitting*, pengurangan data secara geometris pada setiap *node* menyebabkan generalisasi kemampuan yang tidak baik dan data *overfitting* [20]. Berikut adalah ilustrasi dari *Decision Tree*:

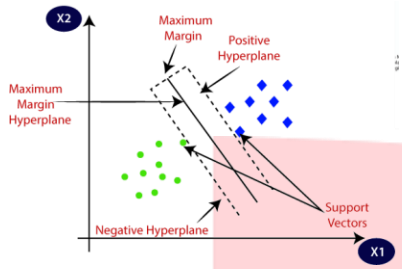


Gambar 2. Ilustrasi *Decision Tree*

E. Support Vector Machnie (SVM)

Support Vector Machine (SVM) merupakan metode *supervised learning* bertujuan untuk mengenal pola dan mengklasifikasi data [21]. SVM menggunakan *hyperplane* (batas keputusan) untuk melakukan proses klasifikasi. SVM akan menentukan *hyperplane* untuk memisahkan dua kelas yang berbeda [22]. SVM sudah terbukti sebagai algoritma pembelajaran untuk kategorisasi teks [14]. Setiap data direpresentasikan sebagai titik yang

ditempatkanpada ruang berdimensi *n* (jumlah fitur pada data) yang kemudian akan dipisahkan secara linear. Algoritma SVM dapat memperoleh probabilitas untuk menemukan jawaban dengan mengembangkan ruang informasi prinsip keruang fitur dimensional tinggi, dimana sebuah *hyperplane* dapat ditemukan [23]. Berikut ilustrasi dari *hyperplane:*



Gambar 3. Ilustrasi *hyperplane* [29]

F. Logistic Regression

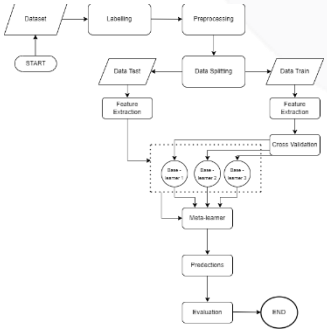
Logistic Regression merupakan algoritma yang digunakan untuk menyelesaikan masalah klasifikasi, dan berbasis dari konsep probabilitas. *Logistic regression* dapat mengklasifikasikan sentimen ke dalam dua kelas dengan label positif dan negatif. *Logistic Regression* memasukkan sebuah hipotesa ke dalam dataset yang merupakan bentuk fungsi matematikal Sigmoid yang konkret [24]. Proses klasifikasi dengan *Logistic Regression*bekerja dengan mengambil fitur bernilai ril dari input, mengalikan masing-masing input sesuai dengan bobot, menjumlahkannya, lalu dikalkulasikan dengan fungsi Sigmoid untuk menghasilkan sebuah probabilitas [25]. Berikut merupakan rumus dari fungsi Sigmoid [26]:

$$0 \leq h\theta\ (z) \leq 1$$

$$Sigmoid\ Function\ \sigma(z) = \frac{1}{1+e^{-z}}$$

III. METODE

A. Flowchart



Gambar 4. Flowchart

Stacking Ensemble akan diterapkan pada sistem yang dibangun dengan mempertimbangkan beberapa kondisi berbeda yang berhubungan dengan proses *pre-processing*, *feature extraction*, dan model-model klasifikasi. Pada penelitian ini, model-model klasifikasi Multinomial Naïve Bayes, *Decision Tree*, dan *Support Vector Machine* dijadikan sebagai *base-learners*, sedangkan model klasifikasi *Logistic Regression* dijadikan sebagai *meta-learner*.Alur sistem dapat dilihat pada gambar di atas.

B. Dataset

Dataset yang digunakan pada peneliatian adalah sebagai berikut ini:

1. 3.000 *tweets* yang diterjemah dari Bahasa Indonesia ke Bahasa Inggris dan diberi label secara otomatisdengan menggunakan model VADER dari *libra*ry NLTK
2. 3.000 *tweets* yang diberi label secara manual dan menggunakan *libra*ry Sastrawi
3. *Tweets* berkaitan dengan informasi dan opini penanganan COVID-19 di Indonesia
4. *Tweets* yang diunggah pada bulan Januari 2021 hingga bulan Desember 2021.

C. *Pre-processing*

Untuk memperoleh label analisis sentimen pada *tweets*, ada beberapa tahapan yang harus dilakukan di antaranya *data cleaning, data pre-processing, data splitting, feature extraction,* klasifikasi dan evaluasi. Pada proses *pre-processing,* dokumen berbentuk teks diubah ke format yang sudah *clean.* Adapun tahapan-tahapan dari*pre-processing* adalah sebagai berikut:

3.3.1 *Tokenizing*

Tokenizing dilakukan dengan cara mengubah dokumen teks menjadi sebuah string dan membaginya ke dalam *list* kata-kata. Contoh dari *tokenizing* bisa dilihat di Tabel 1.

Table 1. *Tokenizing*

Input Tweets	Output
'vaksin yang ada di indonesia'	'vaksin', 'yang', 'ada', 'di', 'indonesia'
'virus corona menyebarluas'	'virus', 'corona', 'menyebarluas'

3.3.1 *Stop Word Removal*

Stop Word Removal merupakan proses penghapusan kata-kata yang tidak memiliki arti signifikan yang sering muncul pada sebuah dokumen teks, dengan tujuan untuk memperoleh informasi dengan mangkus dan mempersingkat alur dari sistem. Kamus yang dijadikan sebagai refensi kata-kata diambil dari *Natural Language Toolkit* (NLTK) versi 3.7. Contoh dari *stop word removal* dapat dilihat di Tabel 2.

Table 2. *Stop Word Removal*

Input Tweets	Output
'saya membenci virus covid'	'membenci', 'virus', 'covid'
'angka covid ternyata malah naik'	'angka', 'covid', 'naik'

3.3.1 *Stemming*

Stemming merupakan proses penguraian dari suata kata menjadi bentuk kata dasarnya dengan cara menghilangkan imbuhan yang bertujuan untuk menyederhanakan proses ekstraksi fitur. Contoh dari *stemming* dapat dilihat di Tabel 3.

Table 3. *Stemming*

Input	Output
'membuat', 'dibuat', 'buatan'	'buat'
'meneliti','penelitian','diteliti'	'teliti'

D. Feature Extraction

Feature Extraction merupakan proses pengubahan data yang sudah melewati proses *pre-processing* menjadidata yang divektorisasi. Peneliti menggunakan metode *Term-Frequency – Inverse Document Frequency* (TF- IDF), yang bertujuan untuk menghitung bobot setiap kata-kata yang muncul pada suatu dokumen teks. Jika suatu kata sering muncul pada dokumen yang berbeda, maka kata tersebut tidak memiliki hal yang menjadi pembeda, sehingga harus diberi bobot yang lebih kecil dibanding dengan kata yang memiliki frekuensi kemunculan rendah. Rumus dari TF-IDF adalah sebagai berikut:

$$w_{i,j=tf_{i,j} x \log \frac{N}{df_i}}$$

Persamaan di atas menunjukkan perhitungan bobot (*w*) dari sebuah kata (*i*) yang muncul dalam suatudokumen (*j*). Frekuensi (*i*) pada dokumen (*j*) ditandai dengan (*tfij*). Jumlah total dokumen ditandai dengan (*N*) dan jumlah dokumen yang mengandung kata (*i*) ditandai dengan (*dfi*).

E. Classification

Pada *Stacking Ensemble,* algoritma *meta-classifier* akan melakukan proses *learning* untuk mengkombinasi prediksi-prediksi yang dihasilkan oleh *base-classifier*s. Keuntungan dari *Stacking Ensemble* adalah bahwa metode ini dapat memanfaatkan kemampuan berbagai macam model-model klasifikasi dan/atau regresi yang baik dan menghasilkan prediksi yang lebih baik dari model-model individu lainnya. Berikut *classifiers* yang penulis gunakan pada penelitian ini:

3.5.1 Base-classifiers

Pada penelitian ini, tiga *base-classifiers* digunakan untuk menentukan sentimen. Mengacu pada penelitian-penelitian sebelumnya yang meneliti analisis sentimen secara luas, peneliti memilih tiga model klasifikasi untuk dijadikan *base-classifiers* yaitu model Multinomial Naïve Bayes, *Decision Tree*, dan *Support Vector Machine.*

3.5.2 Meta-classifier

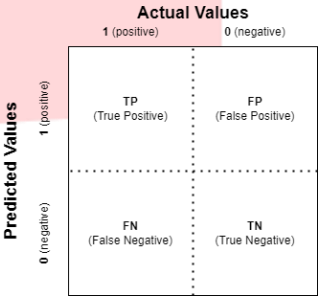
Untuk *meta-classifier* yang digunakan pada penelitian ini, peneliti memilih model *Logistic Regression* yang kemudian akan dilatih berdasarkan prediksi-prediksi yang dibuat oleh tiga *base-learners* pada tahap sebelumnya. Model ini akan menjadikan prediksi-prediksi tersebut sebagai input, dan menghasilkan prediksi akhir. Selain itu, model ini dipilih karena dapat menyelesaikan masalah klasifikasi (pelabelan).

F. Evaluasi Model

Untuk menguji unjuk kerja dari masing-masing model, proses evaluasi harus dilakukan agar dapat mengetahui kecocokkan model dengan data yang digunakan. Evaluasi tersebut dapat diukur dengan menggunakan metrik *Accuracy*, dan dibantu oleh *Confusion Matrix.* Rumus dan ilustrasi adalah sebagai berikut:

3.6.1 *Accuracy* $= \frac{TP + TN}{n}$

3.6.2 *Confusion Matrix*



Gambar 5. Ilustrasi *Confusion Matrix*

Nilai TP (*True Positive)* merupakan nilai pada saat *classifier* mengidentifikasi sebuah data teksyang positif secara akurat dan memasukkannya ke dalam kategori positif. Nilai TN (*True Negative)*merupakan nilai pada saat *classifier* mengidentifikasi sebuah data teks negatif secara akurat dan memasukannya kedalam kategori negatif. Nilai FP (*False Positive)* merupakan nilai pada saat *classifier* mengidentifikasi sebuah data teks negatif secara tidak akurat dan memasukannya ke dalam kategori positif. Nilai FN (*False Negative)* merupakan nilai pada saat *classifier* mengidentifikasi sebuah data teks positif secara tidak akurat dan memasukkannya ke dalam kategori negatif.

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini, terdapat dua skenario yang diujikan:

A. Skenario Pertama

4.1.2 Membandingkan hasil dan akurasi performa dari model *Stacking Ensemble* dengan model-model individu *classifier* lainnya (Multinomial Naïve Bayes, *Decision Tree, dan Support Vector Machine*) menggunakan dataset berupa *tweets* sejumlah 3.000 *tweets.*

4.1.3 *Tweets* tersebut diterjemahkan dari bahasa Indonesia ke bahasa Inggris menggunakan *library* 'googletrans' dan diberi label secara otomatis memanfaatkan model VADER dari *library* 'NLTK(*Natural Language Toolkit*)'.

B. Skenario Kedua

4.1.1 Membandingkan hasil dan akurasi performa dari model *Stacking Ensemble* dengan model-model individu *classifier* lainnya (Multinomial Naïve Bayes, *Decision Tree*, dan *Support Vector Machine*), dengan menggunakan dataset berupa *tweets* sejumlah 3.000 *tweets*.

4.1.2 *Tweets* tersebut label secara manual. Untuk memaksimalkan akurasi secara tatanan bahasa, pada proses *stemming*, *tweets* tersebut diolah menggunakan *library* Sastrawi.

2. Hasil

Kedua skenario yang dijabarkan di atas memiliki rasio yang sama dalam pembagian data *train* dan data *test*, yaitu 80:20. Hasil dari penelitian ini dijelaskan pada tabel sebagai berikut:

Table 4. Persebaran data *tweet* dan presentase sentimennya

	Jumlah Tweet Positif	**Jumlah Tweet Negatif**	**Presentase Sentimen Positif**	**Presentase Sentimen Negatif**
Skenario Pertama	2259	742	75.3%	24.7%
Skenario Kedua	1829	1172	60.69%	39.31%

Pada Tabel 4, ditunjukkan bahwa sentimen pengguna Twitter di Indonesia terhadap penanganan COVID-19 di Indonesia adalah positif, dengan presentasi sentimen positif 75.3% pada skenario pertama, dan 60.69% pada skenario kedua. Dari uraian tersebut dapat ditemukan bahwa penanganan COVID-19 pemerintah Indonesia secara umum mendapatkan reaksi positif dari masyarakat Indonesia.

Table 5.Hasil akurasi skenario pertama

Model	**Accuracy**
Naïve Bayes	76.37%
Decision Tree	76.04%
Support Vector Machine	84.03%
Stacking Ensemble Classifier	84.69%

Pada Tabel 5, ditunjukkan bahwa model *Stacking Ensemble* dapat meningkatkan nilai akurasi sebesar 0.63% dari nilai tertinggi yang dimiliki oleh *base-classifiers*. Di bawah ini merupakan hasil dari *confusion matrix* pada skenario pertama:

Table 6. Hasil *Confusion Matrix* skenario pertama

Model	**Accuracy**
Naïve Bayes	76.48%
Decision Tree	75.85%
Support Vector Machine	83.33%
Stacking Ensemble Classifier	84.35%

Model	**TP**	**FP**	**FN**	**TN**
Naïve Bayes	453	0	142	6
Decision Tree	384	69	75	73
Support Vector Machine	436	17	79	69
Stacking Ensemble Classifier	432	21	71	77

Pada Tabel 6, ditunjukkan bahwa *Stacking Ensemble* memprediksi sejumlah 432 dokumen positif secara akurat, 21 dokumen negatif secara akurat, 71 dokumen positif secara tidak akurat, serta 77 dokumen negatif secara tidak akurat. Dari uraian tersebut, ditemukan bahwa *Stacking Ensemble* dapat meningkatkan performa proses klasifikasi teks pada skenario ini.

Table 7. Hasil akurasi skenario kedua

Pada Tabel 7, ditunjukkan bahwa model *Stacking Ensemble* dapat meningkatkan nilai akurasi sebesar 1.02% dari nilai tertinggi yang dimiliki oleh *base-classifiers*. Di bawah ini merupakan hasil dari *confusion matrix* pada skenario kedua:

Table 8. Hasil *Confusion Matrix* skenario kedua

Model	TP	FP	FN	TN
Naïve Bayes	352	18	85	133
Decision Tree	294	76	66	152
Support Vector Machine	335	35	63	155
Stacking Ensemble Classifier	341	29	63	155

Pada Tabel 8, ditunjukkan bahwa *Stacking Ensemble* memprediksi sejumlah 341 dokumen positif secara akurat, 155 dokumen negatif secara akurat, 29 dokumen positif secara tidak akurat, serta 63 dokumen negatif secara tidak akurat. Dari uraian tersebut, ditemukan bahwa *Stacking Ensemble* dapat meningkatkan performa proses klasifikasi teks pada skenario ini.

V. KESIMPULAN

Berdasarkan evaluasi yang telah dilakukan, ditemukan bahwa reaksi umum pengguna Twitter di Indonesia terhadap penanganan COVID-19 di Indonesia secara umum adalah positif, dengan presentase sentimen positif 75.3% dan 60.39% pada dua skenario yang berbeda. Sehingga, dapat disimpulkan bahwa penanganan COVID-19 di Indonesia dianggap baik oleh masyarakat Indonesia. Selain itu, dapat disimpulkan bahwa metode *hybrid Stacking Ensemble* dapat meningkatkan nilai akurasi yang dihasilkan oleh *classifiers* individu lainnya, dengan perbedaan 0.63% dan 1.02%.

Hal yang perlu dicatat adalah bahwa *Stacking Ensemble* tidak menjamin hasil yang lebih baik dari pada model klasifikasi yang lain, mengingat ada beberapa hal yang menjadi pertimbangan, seperti ketidakseimbangan dataset, perbedaan jenis data, perbedaan fitur, dan lain sebagainya. Peneliti berharap penelitian ini dapat menjadi referensi sebagai pengembangan dan bantuan bagi penilitian analisis sentimen yang menggunakan model *Stacking Ensemble* di masa depan. Saran dari peniliti bagi penelitian yang lebih mendalam adalah untuk menggunakan *base-classifers* yang lebih heterogen serta untuk menambahkan metode seleksi fitur untuk menyelesaikan masalahanalisis sentimen.

REFERENSI

[1] H. Tankovska, "Statista," 9 February 2021. [Online]. Available: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. [Accessed 2 May 2021].

[2] WHO, "WHO," 1 May 2021. [Online]. Available: https://www.who.int/countries/idn/. [Accessed 2 May 2021].

[4] L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," in *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020.

[3] Z. T. Soomro, S. H. W. Ilyas and U. Yaqub, "Sentiment, Count during COVID-19 Pandemic," in *2020 7th International Confe Computing (BESC)*, Bournemouth, United Kingdom, 2020.

[13] V. M., J. Vala and a. P. Balani, "A Survey on Sentiment Analys *ComputerApplications*, vol. 133, no. 9, pp. 7-11, 2016.

[14] A. M. Rahat, A. Kahir and A. K. M. Masum, "Comparison of N Sentiment Analysis Using Review Dataset," in *2019 8th Intern Advancement in Research Trends (SMART)*, Moradabad, 2019.

[15] N. K. Othman, M. Hussin and R. A. R. Mahmood, "Sentiment *International Journal of Engineering and Advanced Technolog* 2019.

[16] H. Yousef, A. &. Medhat, W. &. Mohamed and Hoda., "Sentim A Survey," *Ain Shams Engineering Journal*, vol. 4, pp. 1093-1

[17] A. Mccallum and K. Nigam, "A Comparison of Event Models f on Learning for Text Categorization, Madison, 2001.

[18] S. S. Nazrul, "Multinomial Naïve Bayes Classifier for Text," T https://towardsdatascience.com/multinomial-naive-bayes-classi 8dd6825ece67.

[19] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Temp Mining," *Jurnal Edukasi dan Penelitian Informatika*, vol. V, no

[20] D. Ignatov and A. Ignatov, "Decision Stream: Cultivating Deep *International Conference on Tools with Artificial Intelligence (*

[21] S. K. Lidya, O. S. Sitompul and a. S. Efendi, "Sentiment Analy Menggunakan Support Vector Machine (SVM)," Semin. Nas. T

[22] U. Makhmudah, S. Bukhori, J. A. Putra and a. B. A. B. Yudha, Indonesian Homosexual Tweets Using Support Vector Machin Conference on Computer Science, Information Technology, an (ICOMITEE), Jember, 2019.

[23] S. Rana and A. Singh, "Comparative analysis of sentiment orie Bayes techniques," 2nd International Conference on Next Gene (NGCT), pp. 106- 111, 2016.

[24] S. Tripathi, R. Mehrotra, V. Bansal and S. Upadhyay, "Analyzi Dataset," in 12th International Conference on Computational In Networks (CICN), Bhimtal, 2020.

[25] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of C TF-IDF And Logistic Regresion," in 6th Information Technolo Surabaya, 2020 .

[26] A. Poornima and K. S. Priya, "A Comparative Sentiment Analy Using Machine Learning Techniques," in 6th International Con and Communication Systems (ICACCS), Coimbatore, 2020 .

[27] "Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasi Available: https://covid19.go.id/peta-sebaran-covid19. [Access

[28] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tem Berbasis Text Mining," Jurnal Edukasi dan Penelitian Informat

[29] I. C. Setia, "Support Vector Machine : SVM Implementation us https://insancs.medium.com/support-vector-machine-svm-impl 4442e9a5babc.

[5] J. Burrell, "How the machine 'thinks: Understanding opacity in machine learning algorithms.," *Big Data & Society* , p. 3, 2016.

[6] S. Kumar and R. Singh, "Comparative analysis of ensemble classifiers for sentiment analysis and opinion mining.," in *3rd International Conference on Advances in Computing,Communication & Automation (ICACCA)* , Dehradun, India, 2017 .

[7] N. Abaeikoupaei and H. Al Osman, "A Multi-Modal Stacked Ensemble Model for Bipolar Disorder Classification.," *IEEE Transactions on Affective Computing*, p. 1, 2020.

[8] B. Liu, Sentiment Analysis and Opinion Mining, Chicago: University of Illinois: Morgan & Claypool Publisher, 2012.

[9] R. Feldman, "Techniques and Applications for Sentiment Analysis," *Communications of the ACM*, pp. 82-89, 2013.

[10] V. A. Fitri, R. Andreswari and M. A. Hasibuan, "Sentiment Analysis of Social Media Twitter with a caseof anti-LGBT Campaign in Indonesia using Naive Bayes, Decision Tree, and Random Forest Algorithm," *Procedia Computer Science*, vol. 161, pp. 765-772, 2019.

[11] Z. H. Zhou, " Ensemble learning," *Encyclopedia of Biometrics*, p. 270–273, 2009.

[12] N. F. F. da Silva, E. R. Hruschka and J. E. R. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Decision Support*, no. 66, p. 170–179, 2014.