

Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme *Random Forest* dan *BERT*

Sentiment Analysis of Twitter Users on COVID-19 Vaccines in Indonesia using Random Forest and BERT Algorithms

AMIN ELHAN¹, MEDRIA KUSUMA DEWI HARDHIENATA^{1*}, YENI HERDIYENI¹, SONY HARTONO WIJAYA¹, JULIO ADISANTOSO¹

Abstrak

Pandemi Covid-19 mendorong banyak pihak agar mampu beradaptasi dengan kondisi terkini. Salah satu program yang diluncurkan pemerintah agar dapat mengatasi penyebaran Covid-19 adalah dengan menjalankan program vaksinasi. Agar dapat mengetahui animo masyarakat terkait program vaksinasi Covid-19 yang diluncurkan, perlu dilakukan analisis sentimen. Analisis sentimen pada umumnya dilakukan untuk mendapatkan informasi terkini dari korpus yang besar. Tujuan penelitian ini adalah melakukan analisis sentimen pengguna Twitter terhadap vaksinasi Covid-19 di Indonesia menggunakan Algoritme *Random Forest* dan *Bidirectional Encoder Representation from Transformer (BERT)*. Tahapan penelitian yang dilakukan meliputi praproses data Twitter yang berkaitan dengan vaksinasi Covid-19, pelabelan sentimen, penanganan data yang tidak seimbang, pengklasifikasian dataset menggunakan algoritme *Random Forest* dan *BERT*, serta analisis dan evaluasi. Setelah dilakukan penanganan data yang tidak seimbang, hasil analisis sentimen pengguna *Twitter* terhadap vaksinasi Covid-19 di Indonesia menghasilkan akurasi sebesar 81%, *F1-score* sebesar 74%, presisi 76%, dan *recall* 74%. menggunakan algoritme *Random Forest* dan akurasi sebesar 82%, *F1-score* 79%, presisi 78%, dan *recall* 79%. menggunakan Algoritme *BERT*. Meskipun secara umum Algoritme *BERT* memiliki kinerja klasifikasi yang sedikit lebih baik dari Algoritme *Random Forest*, hasil simulasi menunjukkan bahwa algoritme *Random Forest* memiliki waktu komputasi yang signifikan lebih rendah dibandingkan dengan algoritme *BERT* pada kasus yang diujikan.

Kata Kunci: analisis sentimen, Covid-19, vaksinasi, *Random Forest*, *BERT*, Twitter

Abstract

The Covid-19 pandemic has encouraged many stakeholders to be able to adapt to current conditions. One of the programs launched by the government in order to overcome the spread of Covid-19 is to run a vaccination program. In order to find out the public's interest in the Covid-19 vaccination program that was launched, it is necessary to carry out a sentiment analysis. Sentiment analysis is generally done to obtain the latest information from a large corpus. The purpose of this study is to analyze the sentiments of Twitter users towards the Covid-19 vaccination in Indonesia using the Random Forest and BERT Algorithms. The research stages include pre-processing Twitter data related to Covid-19 vaccination topics, sentiment labeling, handling unbalanced data, classifying datasets using the Random Forest and BERT algorithms, as well as analysis and evaluation. After handling unbalanced data, the results of Twitter user sentiment analysis for Covid-19 vaccination in Indonesia yielded an accuracy of 81%, F1-score of 74%, precision of 76%, and recall of 74% using the Random Forest algorithm and an accuracy of 82%, F1-score 79%, precision of 78%,

¹Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor 16680
*Penulis Korespondensi. Surel: medria.hardhienata@apps.ipb.ac.id

and recall of 79% using the BERT Algorithm. Although the BERT Algorithm has generally a slightly higher performance than the Random Forest Algorithm, the simulation results show that the Random Forest algorithm has significantly lower computation time compared to the BERT algorithm in the considered case.

Keywords: sentiment analysis, Covid-19, vaccination, Random Forest, BERT, Twitter

PENDAHULUAN

Tipe baru *Coronavirus* dengan tipe SARS-CoV-2 yang menyebabkan epidemi penyakit pernapasan akut muncul pada bulan Desember 2019 di kota Wuhan, China (Bavel *et al.* 2020). The *World Health Organization* (WHO) mendeklarasikan pandemi Covid-19 sejak 11 Maret 2020, dengan bukti *outbreak* penyakit Covid-19 yang terjadi di beberapa negara seperti Korea Utara, Iran, Amerika Serikat, dan Eropa (Rockett *et al.* 2020). Kebutuhan informasi saat ini terkait Covid-19 perlu didukung dengan teknologi canggih yang mampu diakses secara luas. Salah satu *platform* sosial media, yaitu Twitter mengalami peningkatan aktivitas yang cukup signifikan saat pandemi Covid-19 (Rosenberg *et al.* 2020). Selama pandemi Covid-19, pemerintah di berbagai negara menggunakan Twitter sebagai kanal komunikasi untuk mengabarkan kebijakan terbaru dan berita terkait Covid-19 ke ranah publik (Xue *et al.* 2020).

Media *microblogging* Twitter merupakan alat komunikasi yang dapat membagikan informasi mengenai opini seseorang, dengan format bebas dan akses yang sangat mudah (Pak dan Paroubek 2020). Jumlah pengguna Twitter saat ini sudah mencapai lebih dari 3,8 miliar orang (Cuello-Garcia *et al.* 2020). Penggunaan Twitter sebagai data analisis untuk melakukan prediksi sudah pernah dipakai sebelumnya dalam berbagai bidang termasuk dalam bidang politik seperti dalam penyelenggaraan pemilihan umum di Jerman, Singapura, Irlandia, Amerika Serikat, dan Prancis (Qamar *et al.* 2014). Selain dalam bidang politik, berbagai bentuk analisis dengan data Twitter dilakukan oleh para peneliti di dunia untuk memahami perubahan aktivitas selama pandemi Covid-19 (Barkur *et al.* 2020), (Sastramidjaja dan Rosli 2021). Analisis sentimen data Twitter terkait kebijakan *lockdown* di India, contohnya, memperlihatkan sentimen positif masyarakat dan negara tersebut dan pemahaman masyarakat tersebut terhadap kebijakan pemerintah berakibat pada menurunnya jumlah penderita Covid-19 (Barkur *et al.* 2020). Di Indonesia sendiri terdapat beberapa metode penanganan kasus covid, salah satunya dengan diluncurkannya program vaksinasi. Vaksinasi pertama Covid-19 di Indonesia dimulai pada 13 Januari 2021 yang diawali oleh Presiden Indonesia Joko Widodo (Sastramidjaja dan Rosli 2021). Hal ini tentunya menimbulkan berbagai reaksi dari masyarakat pada berbagai platform sosial media, terutama Twitter, karena perbedaan pandangan terhadap vaksinasi.

Berbagai bentuk analisis sentimen berupa klasifikasi *tweet* di Twitter telah dilakukan di Indonesia, seperti analisis dengan kata kunci “vaksin merah putih” dan “*sinovac*” (Laurensz *et al.* 2021). Penelitian tersebut menggunakan 845 *tweet* dan menggunakan *classifier* SVM dan Naïve Bayes untuk menganalisis sentimen pengguna Twitter. Hasil penelitian tersebut menunjukkan bahwa dihasilkan akurasi sebesar 84% dengan metode SVM dan *classifier* Naïve Bayes sebesar 85% (Laurensz *et al.* 2021). Analisis menggunakan data *Twitter* juga telah dilakukan dengan metode klasifikasi *Support Vector Machine* (SVM) serta menghasilkan nilai akurasi sebesar 90% (Chairunnisa *et al.* 2022). Meski demikian, pada penelitian Chairunnisa *et al.* 2022 belum dilakukan proses penghilangan duplikasi data.

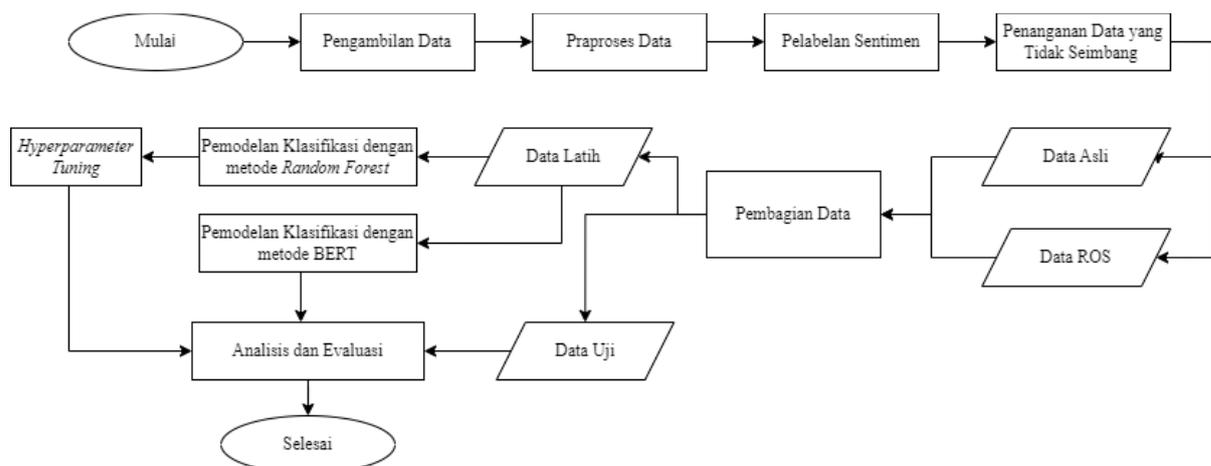
Penelitian lain juga telah dilakukan dengan metode *Random Forest* yang terbukti memberikan hasil akurasi yang cukup baik untuk sentimen analisis (Soumya dan Pramod, 2020). Berdasarkan penelitian tersebut, metode *Random Forest* terbukti memberikan akurasi yang baik sebesar 95% dalam analisis Bahasa *Malayam* (Bahasa lokal India). Selain

metode *Random Forest*, juga terdapat metode *Bidirectional Encoder Representation from Transformer* (BERT) yang digunakan untuk melakukan sentimen analisis. Penggunaan Metode BERT dengan *dataset* Semeval-2014 menghasilkan nilai presisi sebesar 92.78% (Li *et al.* 2020).

Meski berbagai metode analisis sentimen telah dilakukan untuk melihat sentimen masyarakat terhadap kebijakan vaksinasi pemerintah di beberapa negara, belum banyak penelitian yang dilakukan untuk melihat sentimen masyarakat terkait kebijakan vaksinasi Covid-19 di Indonesia dengan pendekatan *Random Forest* dan BERT. Dengan latar belakang tersebut, penelitian ini bertujuan melakukan analisis sentimen pengguna Twitter terhadap vaksinasi Covid-19 di Indonesia menggunakan Algoritme *Random Forest* dan *BERT*. Tahapan penelitian yang dilakukan meliputi praproses data Twitter yang berkaitan dengan vaksinasi Covid-19, pelabelan sentimen, penanganan data yang tidak seimbang, pengklasifikasian dataset menggunakan algoritme *Random Forest* dan *BERT*, serta analisis dan evaluasi.

METODE

Tahapan penelitian yang dilakukan meliputi pengambilan data, praproses data, pelabelan sentimen, penanganan data yang tidak seimbang, pembagian data menjadi data latih dan data uji, pemodelan klasifikasi dengan metode *Random Forest* dan BERT, *hyperparameter tuning* untuk algoritme *Random Forest*, serta analisis dan evaluasi. Gambar 1 menjelaskan alur dari penelitian yang dilakukan.



Gambar 1 Alur penelitian

Pengambilan Data

Data *tweet* diambil menggunakan *Package snsrape* yang tersedia dalam bahasa pemrograman Python. Data yang diambil sebanyak 47140 *tweet* dengan kata kunci “vaksin corona”. Rentang tanggal *tweet* yang diambil mulai dari 1 September 2020 hingga 30 Juni 2021. Penelitian ini menggunakan data awal yang sama dengan penelitian Chairunnisa *et al.* (2022) yakni sebanyak 42163 *tweet* (Chairunnisa *et al.* 2022). Namun, berbeda dengan penelitian Chairunnisa *et al.* (2022), data *tweet* dalam penelitian ini dibersihkan terlebih dahulu dari *tweet* duplikat sehingga tersisa 18805 *tweet*.

Praproses Data

Tahap praproses data dilakukan agar terjadi standarisasi bentuk kata, memperbaiki kata yang kurang sempurna, menghilangkan URL twitter yang masih terkandung dalam *cell*, dan objek lainnya (Angiani *et al.* 2016). Berikut praproses yang dilakukan dalam penelitian ini:

1. *Case Folding*

Merupakan proses untuk menyamakan seluruh bentuk kata. Huruf kapital diproses menjadi huruf kecil. Proses ini bertujuan mengurangi indeks berlebih pada kata yang sama.

2. *Data Cleaning*

Proses pembersihan bertujuan mengurangi *noise* pada data. Proses ini menghapus simbol, tanda baca, angka, tautan URL, emotikon, *mention*, *hashtag*, dan *stopword*.

3. Normalisasi Bahasa

Mengembalikan bentuk penulisan yang tidak baku menjadi penulisan yang sesuai KBBI seperti “gk” menjadi “tidak”, dan kata-kata singkatan lainnya.

4. Tokenisasi

Proses pemecahan kalimat menjadi potongan atau *token* untuk diproses lebih lanjut dengan program. Berikut contoh dari bentuk tokenisasi: ['ndak', 'usah', 'panikan', 'pak', 'tetap', 'patuhi', 'prokes'].

5. Penghapusan *Stopword*

Kata-kata yang sering muncul namun tidak mempunyai pemaknaan dalam proses ini. Contoh *stopword* yang dimaksud yaitu: “yang”, “dan”, “di”.

6. *Stemming*

Proses ini mengubah kata menjadi bentuk dasarnya, sehingga tidak ada pemaknaan lebih dari bentuk kata yang sama. Contohnya “memahami” diganti dengan kata “paham”. Package *sastrawi* digunakan dalam proses ini untuk membantu proses *stemming* dengan Bahasa Indonesia. Beberapa data *tweet* yang sudah diproses terlihat menjadi seperti ini:

['saya', 'pikir', 'ini', 'konsep', 'sangat', 'bagus', 'kebijakan', 'pemerintah', 'terkait', 'akan', 'memberikan', 'vaksin', 'corona', 'virus', 'disease', 'covid', 'secara', 'gratis', 'diutamakan', 'bagi', 'masyarakat', 'tidak', 'mampu', 'menjadi', 'bukti', 'kehadiran', 'negara'].

['tim', 'riset', 'uji', 'klinis', 'vaksin', 'covid', 'buatan', 'sinovac', 'membenarkan', 'adanya', 'relawan', 'positif', 'terpapar', 'virus', 'corona'].

Pelabelan Sentimen

Pelabelan dibagi menjadi tiga kelas, yaitu positif, negatif, dan netral. Sentimen positif menunjukkan setuju atau dukungan terhadap vaksin dan kebijakan pemerintah. Sentimen negatif berisi pernyataan kontra terhadap vaksin dan kebijakan pemerintah. Sentimen netral berisi berita atau informasi yang tidak mengarah kepada dukungan maupun kontra. Pelabelan manual dilakukan oleh 13 orang mahasiswa IPB *University* dari berbagai program studi dengan panduan pelabelan yang telah disepakati. Hasil pelabelan manual menunjukkan jumlah sentimen positif sebesar 5841 *tweet*, sentimen negatif sebesar 3256 *tweet*, dan sentimen netral sebesar 9708 *tweet*.

Penanganan Data Tidak Seimbang

Penanganan data tidak seimbang dilakukan dengan metode *Random Oversampling* (ROS) (Zheng *et al.* 2015). Metode ROS merupakan metode untuk menyeimbangkan distribusi kelas dengan melakukan replikasi dari kelas minoritas.

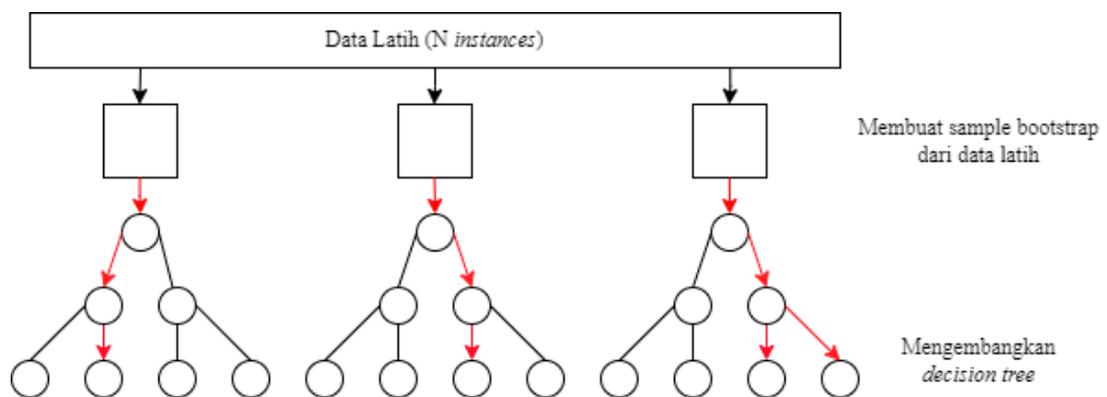
Analisis Sentimen dengan *Random Forest*

Random forest merupakan kombinasi dari pohon prediksi dimana setiap pohon bergantung pada nilai vektor acak (Breiman 2001). Algoritme ini menggabungkan konsep dari *random subspaces* dan *bootstrap aggregating (bagging)*. Algoritme *Random Forest* dapat dijabarkan sebagai berikut. Pertama, dilakukan *bootstrap sampling* dengan penggantian dari data *train*. Kedua, untuk setiap *bootstrap* pada tahap pertama, pohon akan

berkembang dengan *sampling* acak dari variabel input dan pilih pembagian terbaik dari variabel tersebut. Kemudian tahap pertama dan kedua diulang hingga iterasi ke- k hingga pohon terbentuk. Setelah dilakukan *training*, prediksi dilakukan dengan mengambil nilai rata-rata dari semua pohon regresi. Persamaan regresi dijabarkan sebagai berikut (al Amrani *et al.* 2018):

$$F_{rf}^B = \frac{1}{B} \sum_{b=1}^B T_b(x), \quad (1)$$

dengan B merupakan jumlah ulangan *bagging*, dan T_b merupakan nilai dari pohon regresi. Untuk klasifikasi, prediksi diambil dari *majority voting* dari setiap pohon klasifikasi. Gambar 2 menjelaskan bagaimana metode *Random Forest* mengambil kesimpulan menggunakan *majority voting* (al Amrani *et al.* 2018).



Gambar 2 Representasi *Random Forest*

Analisis sentimen menggunakan Algoritme *Random Forest* dilakukan untuk mendapatkan kinerja klasifikasi model dalam memilih kategori label dari data yang sudah dilatih. *Random Forest* melakukan proses *bootstrap* untuk menarik contoh acak untuk kemudian dijadikan data latih. Pohon klasifikasi dibuat dengan menggunakan *random feature selection*. Peubah penjelas dipilih yang terbaik untuk dijadikan sebagai penyekat dan dilanjutkan dengan pemisahan menjadi dua simpul baru. Proses ini terus berlanjut hingga ukuran minimum amatan dalam simpul tercapai. Pembuatan pohon terus dilakukan hingga diperoleh pohon klasifikasi. Tiap pohon klasifikasi menghasilkan satu suara, dan penentuan klasifikasi didasarkan pada suara terbanyak (*majority vote*). Package Python yang digunakan dalam penelitian ini untuk membangun model *Random Forest* adalah *RandomForestClassifier* dari *Sklearn*.

Hyperparameter Tuning

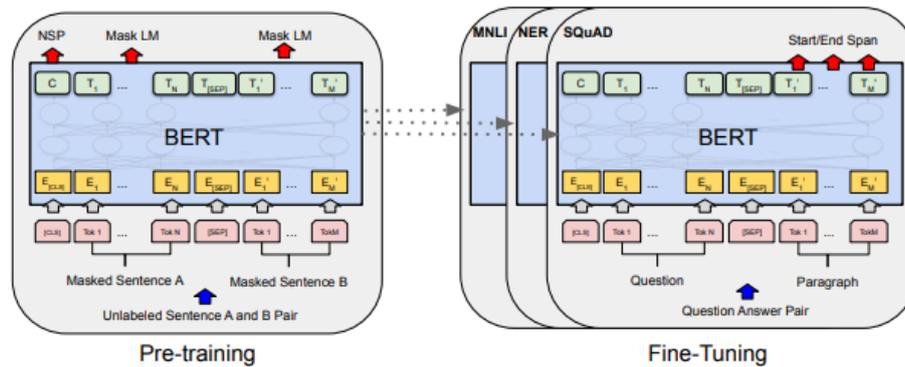
Hyperparameter tuning pada *Random Forest* digunakan dengan cara *grid search*. Metode *grid search* mencoba semua kombinasi dari semua parameter yang digunakan (Probst *et al.* 2018). Ruang pencarian untuk tiap *hyperparameter* adalah *minimum sample leaf* 1-5, *minimum sample split* 2-5 dan *n estimator* 100, 200, 300 dan 1000. *Hyperparameter Tuning* dilakukan dengan menggunakan metode *grid search cross validation* dengan *10-fold cross validation*.

K-Fold Cross Validation

K-fold cross validation merupakan salah satu cara untuk membagi gugus data menjadi data latih dan data uji. Metode ini digunakan untuk mengurangi bias dalam pengambilan contoh. merupakan nilai untuk menentukan banyaknya pembagian gugus data, dimana Sebagian dari gugus data dijadikan data uji pada pemodelan. Penggunaan cross validation berulang direkomendasikan, selama masih bisa dikomputasikan. Penggunaan $k=5$ dan $k=10$ lebih baik, karena bias akan berkurang jika dibandingkan dengan $k=2$ (Rodríguez *et al.* 2010).

Bidirectional Encoder Representation from Transformer (BERT)

BERT merupakan singkatan dari *Bidirectional Encoder Representation from Transformer* (Devlin *et al.* 2018). BERT didesain untuk melakukan *pre-training* dengan representasi *deep bidirectional* dari teks tidak berlabel (Devlin *et al.* 2018). *Encoder* pada BERT membaca semua *sequence* dari kata yang masuk secara menyeluruh. Model dibuat untuk dapat mengerti makna yang terkandung sebelum dan sesudah kata. Dikarenakan kemampuan BERT untuk memahami makna, BERT dapat meningkatkan performa pada berbagai kegiatan pada bidang NLP seperti *Natural Language Inference (NLI)* dan *Question Answering (QA)* (Azhar dan Khodra 2020).



Gambar 3 Ilustrasi *pre-training* dan *fine-tuning* BERT (Devlin *et al.* 2018)

Gambar 3 menunjukkan prosedur *pre-training* dan *fine-tuning* yang dilakukan dengan metode BERT. Arsitektur yang sama digunakan pada *pre-training* dan *fine-tuning*, perbedaan terletak pada *output layer*. *Pre-trained* model digunakan untuk menginisiasi model untuk pekerjaan *downstream* yang berbeda. Saat *fine-tuning*, semua parameter sudah diatur dengan baik (Devlin *et al.* 2018).

Analisis dan Evaluasi Model Klasifikasi

Tahap analisis dan evaluasi dilakukan untuk menilai kinerja model yang dibangun. Kinerja algoritme diukur dengan menggunakan *confusion matrix* dan mengukur nilai akurasi, presisi, *recall*, dan *F1-score*, dan *support*.

Confusion Matrix

Klasifikasi menentukan prediksi akan memberikan empat hasil, diantaranya *confusion matrix*, akurasi, presisi, *recall*, dan *F1-score*. *Confusion Matrix* untuk data tiga kelas dijelaskan pada Tabel 1 (Manning *et al.* 2008).

Tabel 1 *Confusion Matrix*

		Prediksi	
		Negatif	Netral
Fakta	Positif		
Positif	TP	FNg1	FNt1
Negatif	FP1	TNg	FNt2
Netral	FP2	FNg2	TNt

Pada Tabel 1, TP adalah jumlah prediksi yang benar untuk data aktual positif, FP1 dan FP2 adalah jumlah prediksi yang salah untuk data aktual positif, TNt adalah jumlah prediksi yang benar untuk data aktual netral, FNt1 dan FNt2 adalah jumlah prediksi yang salah untuk data netral, TNg adalah jumlah prediksi yang benar untuk data aktual negatif, FNg1 dan FNg2 adalah jumlah prediksi yang salah untuk data aktual negatif.

Perhitungan akurasi, presisi, *recall* ditampilkan di bawah ini (Sholehah 2018):

$$Akurasi = \frac{(TP+TNt + TNg)}{(TP + FNg1 + FNt1 + FP1 + TNg + FNt2 + FP2 + FNg2 + TNt)}, \quad (2)$$

$$Presisi = \frac{TP}{TP+FP1+FP2}, \quad (3)$$

$$Recall = \frac{TP}{TP + FNg1 + FNt1}. \quad (4)$$

Sedangkan perhitungan *F1-score* dihitung sebagai berikut (Grandini *et al.* 2020):

$$F1 - score = 2 * \frac{presisi * recall}{presisi + recall}. \quad (5)$$

Pada Persamaan (2), akurasi menunjukkan perbandingan jumlah prediksi yang benar dengan jumlah *input samples*. Presisi pada Persamaan (3) menunjukkan berapa banyak nilai *true positive* di antara semua nilai yang diprediksi positif. *Recall* pada Persamaan (4) menunjukkan berapa nilai *true positive* di antara semua nilai yang aktual positif. *F1-score* dapat diinterpretasikan sebagai rata-rata bobot antara presisi dan *recall* (Grandini *et al.* 2020). Pada *F1-score* merupakan 0 merupakan nilai terburuk dan 1 merupakan nilai terbaik.

Pendekatan *Macro Average* juga digunakan untuk mengukur besarnya rata-rata nilai akurasi, presisi, *recall*, dan *F1-score* dari semua kelas. Nilai *Macro Average* yang besar menandakan algoritme mempunyai kinerja yang baik pada semua kelas, dan sebaliknya (Grandini *et al.* 2020). Perhitungan nilai *Macro Average* dijabarkan sebagai berikut (Grandini *et al.* 2020):

$$\text{MacroAveragePrecision} = \frac{\sum_{k=1}^K \text{Precision}_k}{K}, \quad (6)$$

$$\text{MacroAverageRecall} = \frac{\sum_{k=1}^K \text{Recall}_k}{K}, \quad (7)$$

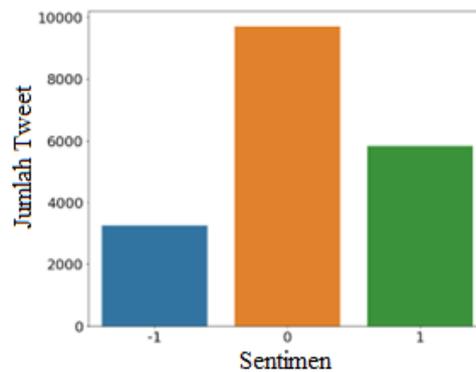
$$\text{Macro F1 - score} = 2 * \frac{\text{MacroAveragePrecision} * \text{MacroAverageRecall}}{\text{MacroAveragePrecision}^{-1} * \text{MacroAverageRecall}^{-1}}, \quad (8)$$

dengan K adalah banyak kelas yang ada pada dataset.

HASIL DAN PEMBAHASAN

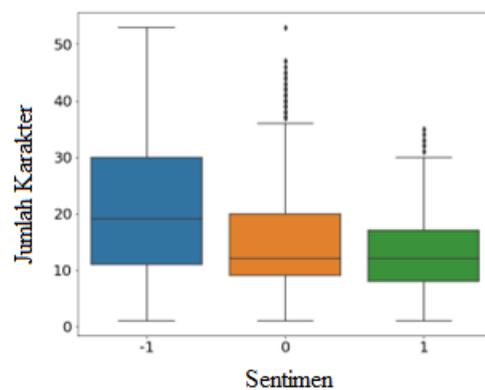
Pengumpulan Data

Data yang dihimpun berjumlah 18805 *tweet* dengan rincian sentimen positif sebesar 5841 *tweet*, sentimen negatif sebesar 3256 *tweet*, dan sentimen netral sebesar 9708 *tweet*. Gambar 4 menunjukkan perbandingan jumlah dokumen atau *tweet* dari setiap kelas sentimen. Perbedaan jumlah *tweet* pada tiap kelas yang cukup signifikan membuat metode penyeimbangan data menggunakan *Random Over Sampling* (ROS) diterapkan pada proses selanjutnya.



Gambar 4 Perbandingan jumlah data *tweet* tiap kelas

Jika dilihat dari panjangnya teks, sentimen negatif terlihat mempunyai panjang teks yang paling panjang di antara dua sentimen lainnya. Sentimen negatif mempunyai rata-rata 19 kata, sentimen netral dengan rata-rata 12 kata, dan sentimen positif dengan rata-rata 12 kata. Gambar 5 menunjukkan perbandingan panjang kata dari setiap kelas.



Gambar 5 Perbandingan panjang kata tiap kelas

Pembagian Data

Data dibagi menjadi data latih dan data uji dengan proporsi 80% dan 20%. Penelitian terkait perbandingan data *train* dan *test* menyarankan rasio 80:20 sebagai pilihan untuk menjadi rasio pemisahan data, hal ini bertujuan memberikan data latih yang cukup untuk klasifikasi multi kelas (Rácz *et al.* 2021). Jumlah pembagian data latih dan data uji dapat dilihat pada Tabel 2.

Tabel 2 Jumlah pembagian data

Sentimen	Data Latih	Data Uji
Positif	4685	648
Netral	7751	1957
Negatif	2608	1156
Total	15044	3761

Hyperparameter Tuning

Hyperparameter Tuning untuk algoritme *Random Forest* dilakukan dengan menggunakan metode *Grid Search Cross Validation* dengan *10-fold cross validation*. Ruang pencarian untuk tiap *hyperparameter* adalah *minimum sample leaf* 1–5, *minimum sample split* 2–5 dan *n estimator* 100, 200, 300 dan 1000. Nilai *tuning* yang didapatkan dari setiap penerapan penanganan asumsi data *imbalance* dapat dilihat pada Tabel 3.

Tabel 3 Parameter terbaik hasil *hyperparameter tuning*

Penanganan Data	Parameter
Data asli	<i>min_samples_split=4, n_estimators=1000</i>
<i>Random Over Sampling</i> (ROS)	<i>n_estimators=300</i>

Evaluasi Model Klasifikasi dengan Data Asli

Hasil klasifikasi menggunakan *Random Forest* yang diukur dengan *confusion matrix* dapat dilihat pada Tabel 4. Berdasarkan Tabel 4, terdapat 148 *tweet* negatif, 448 *tweet* netral, dan 52 *tweet* positif yang diklasifikasikan sebagai negatif. Terdapat 35 *tweet* negatif, 1819 *tweet* netral, dan 103 *tweet* positif yang diklasifikasikan sebagai netral. Terdapat 0 *tweet* negatif, 76 *tweet* netral dan 1080 *tweet* positif yang diklasifikasikan sebagai positif.

Tabel 4 *Confusion Matrix* dari data asli dengan menggunakan metode *Random Forest*

	-1	0	1
-1	148	448	52
0	35	1819	103
1	0	76	1080

Nilai akurasi dan *F1-score* dari model *Random Forest* terlihat menghasilkan sebesar 81% dan 70% dengan data asli. Terlihat nilai *recall* dan *F1-score* pada sentimen negatif terlihat sangat kecil yaitu 23% dan 36%. *Classification Report* dari data asli dapat dilihat pada Tabel 5.

Tabel 5 *Classification Report* dari data asli untuk metode *Random Forest*

	Presisi	Recall	F1-score	Support
-1	0.81	0.23	0.36	648
0	0.78	0.93	0.85	1957
1	0.87	0.93	0.90	1156
Akurasi			0.81	3761
Macro Avg	0.82	0.70	0.70	3761

Selanjutnya dalam penelitian ini juga dilakukan evaluasi model dengan melakukan penanganan dengan data tidak seimbang. Data *tweet* dianggap mempunyai data yang tidak seimbang karena perbandingan antara data *tweet* positif, netral dan negatif adalah 3:5:9. Penanganan data tidak seimbang dilakukan dengan metode *Random Over Sampling* (ROS).

Evaluasi Model Klasifikasi dengan *Random Over Sampling* (ROS)

Random Over Sampling (ROS) menggunakan data penyetaraan jumlah data latih dengan data terbesar dari kelas yang ada (Zheng *et al.* 2015). Pada kasus ini, data kelas netral mempunyai jumlah data terbesar yaitu 7751 *tweet*. Data tiap kelas disetarakan menjadi 7751 *tweet* dengan pilihan acak.

Confusion Matrix dengan menggunakan teknik *Random Forest* dari data ROS dapat dilihat pada Tabel 6. Dengan metode *Random Forest*, nilai akurasi dengan data yang telah diseimbangkan mempunyai nilai yang sama dengan data asli (lihat Tabel 7). Meski demikian, *F1-score* dan *recall* model terlihat lebih tinggi dibanding data asli.

Tabel 6 *Confusion Matrix* dari data ROS untuk metode *Random Forest*

	-1	0	1
-1	270	331	50
0	163	1659	119
1	2	68	1098

Tabel 7 *Classification Report* dari data ROS untuk metode *Random Forest*

	Presisi	Recall	F1-score	Support
-1	0.62	0.41	0.50	651
0	0.81	0.85	0.83	1941
1	0.87	0.94	0.90	1168
Akurasi			0.81	3760
Macro Avg	0.76	0.74	0.74	3760

Perbandingan hasil Klasifikasi antara Metode *Random Forest* dan Metode BERT

Seperti yang telah dijelaskan dalam Bab Metodologi, pada penelitian ini juga dilakukan pembangunan model klasifikasi dengan metode BERT. Tabel 8 dan 9 menunjukkan confusion matrix dari metode BERT dengan data asli dan data yang telah diseimbangkan.

Tabel 8 *Confusion Matrix* klasifikasi BERT data asli

	-1	0	1
-1	371	254	26
0	225	1669	47
1	17	124	1027

Tabel 9 *Confusion Matrix* klasifikasi BERT data ROS

	-1	0	1
-1	391	226	34
0	246	1620	75
1	16	71	1081

Tabel 10 menunjukkan *classification report* untuk Algoritme *Random Forest* dan BERT. Terlihat secara umum, algoritme BERT lebih baik dari Algoritme *Random Forest* dari sisi akurasi, *F1-score* dan *recall*. Sedangkan Algoritme *Random Forest* dengan data asli memiliki kinerja yang lebih baik dari segi presisi. Selain itu algoritme *Random Forest* juga memiliki waktu komputasi yang secara signifikan lebih baik dari algoritme BERT dalam kasus diujikan.

Tabel 10 Perbandingan *Classification Report* dengan Algoritme *Random Forest* dan BERT

Penanganan Data	Akurasi	<i>F1-score</i>	Presisi	<i>Recall</i>	Waktu Komputasi
<i>Random Forest</i> dengan Data Asli	81%	70%	81%	70%	100 menit
<i>Random Forest</i> dengan Data ROS	81%	74%	76%	74%	140 menit
BERT dengan Data Asli	82%	78%	78%	77%	454 menit
BERT dengan Data ROS	82%	79%	78%	79%	759 menit

SIMPULAN

Penelitian ini telah berhasil melakukan sentimen analisis data twitter terkait vaksinasi COVID-19 di Indonesia menggunakan algoritme *Random Forest* dan BERT. Pada penelitian ini dilakukan penanganan data yang tidak seimbang menggunakan teknik *Random Over Sampling* (ROS).

Pengklasifikasian *Random Forest* dengan data asli menghasilkan akurasi 81%, presisi 82%, *recall* 70%, *F1-score* 70%, dan sebesar 3761. Penanganan data dengan *Random Over Sampling* (ROS) untuk algoritme *Random Forest* menghasilkan akurasi 81%, *precision* 76%, *recall* 74%, *F1-score* 74%, dan *support* 3760. Sedangkan pengklasifikasian dengan BERT dengan data asli menghasilkan akurasi 82%, *precision* 78%, *recall* 77%, *F1-score* 78%, dan *support* 3760. Penanganan data dengan *Random Over Sampling* (ROS) untuk algoritme BERT menghasilkan akurasi 82%, *precision* 78%, *recall* 79%, *F1-score* 79%, dan *support* 3760.

Dari hasil penelitian dapat dilihat bahwa pendekatan BERT secara umum memiliki kinerja yang lebih baik dari segi akurasi, *F1-score*, *recall* dibandingkan algoritme *Random Forest*. Meski demikian, waktu komputasi algoritme BERT pada penelitian ini lebih tinggi jika dibandingkan teknik *Random Forest* untuk data yang diujikan.

DAFTAR PUSTAKA

- Al Amrani Y, Lazaar M, el Kadirp KE. 2018. Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*. 127:511–520. doi: 10.1016/j.procs.2018.01.150.
- Angiani G, Ferrari L, Fontanini T, Fornacciari P, Iotti E, Magliani F, Manicardi S. 2016. A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. Di dalam: Armano G, Bozzon A, Cristani M, Giuliani A, editor. Proceedings of the 2nd International Workshop on Knowledge Discovery on the WEB. Knowledge Discovery on the WEB; 2016 Sep 8-10; Cagliari, Italy. Cagliari: hlm 58-69; [diakses 2022 Ags 1]. <https://ceur-ws.org/Vol-1748/paper-06.pdf>.
- Azhar AN, Khodra ML. 2020. Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis. Di dalam: 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA); 2020 Sep 8-9; Aichi, Japan. Aichi: hlm 1-6.
- Barkur G, Vibha, Kamath GB. 2020. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian J Psychiatr*. 51. doi:10.1016/j.ajp.2020.102089.
- Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN, *et al.* 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nat Hum Behav*. 4(5):460–471. doi:10.1038/s41562-020-0884-z.
- Breiman, L. 2001. *Random Forests*. Machine Learning. 45(1):5-32.
- Chairunnisa QA, Herdiyeni Y, Kusuma M, Hardhienata D, Adisantoso J. 2022. Analisis sentimen pengguna Twitter terhadap program vaksinasi Covid-19 di Indonesia menggunakan algoritme Support Vector Machine. *Jurnal Ilmu Komputer & Agri-informatika*. 9(1):79-89. doi:10.29244/jika.9.1.79-89.
- Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. 2020. Social media can have an impact on how we manage and investigate the COVID-19 pandemic. *J Clin Epidemiol*. 127:198–201. doi:10.1016/j.jclinepi.2020.06.028.
- Devlin J, Chang MW, Lee K, Toutanova K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Di dalam: Burstein J, Doran C, Solorio T, editor. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2019 Jun 2-7; Minnesota, USA. Minnesota:Association for Computational Language hlm 4171-4186; <http://arxiv.org/abs/1810.04805>
- Grandini M, Bagli E, Visani G. 2020. Metrics for Multi-Class Classification: an Overview. arXiv. doi : 10.48550/arXiv.2008.05756
- Laurensz B, Sentimen A, Sedyono E. 2021. Analisis sentimen masyarakat terhadap tindakan vaksinasi dalam upaya mengatasi pandemi Covid-19 (analysis of public sentiment on vaccination in efforts to overcome the Covid-19 pandemic). *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*. 10(2):118-123. doi: 10.22146/jnteti.v10i2.1421.
- Li X, Fu X, Xu G, Yang Y, Wang J, Jin L, Liu Q, Xiang T. 2020. Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access*. 8:46868–46876. doi: 10.1109/ACCESS.2020.2978511.
- Manning, Christopher D., Raghavan, Prabhakar, dan Schütze, Hinrich. 2008. An Introduction to Information Retrieval. Cambridge University Press Cambridge, England.

- Pak A, Paroubek P. 2020. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Di dalam: Calzolari N, Choukri K, Maegaard B. Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010; 2010 May 17-23; Valletta, Malta. Malta: hlm 1320-1326.
- Probst P, Wright M, Boulesteix AL. 2018. Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowl Discov*. doi:10.1002/widm.1301.
- Qamar AM, Razzaq MA, Bilal M. 2014. Prediction and Analysis of Pakistan Election 2013 based on Sentiment Analysis. Di dalam: Wu X, Ester M, Xu G, editor. ASONAM '14: Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2014 Agu 17-20; Beijing, China. Beijing: hlm 700-703; [diakses 2022 Jul 3]. doi: 10.1109/ASONAM.2014.6921662.
- Rácz A, Bajusz D, Héberger K. 2021. Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification. *Molecules*. 26(4). doi:10.3390/molecules26041111.
- Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray KA, Eden JS, Chang S, Gall M, Draper J, *et al*. 2020. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat Med*. 26(9):1398–1404. doi:10.1038/s41591-020-1000-7.
- Rodríguez JD, Pérez A, Lozano JA. 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 32(3):569–575. doi:10.1109/TPAMI.2009.187
- Rosenberg H, Syed S, Rezaie S. 2020. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine*. 22(4):418–421. doi:10.1017/cem.2020.361.
- Sastramidjaja Y, Rosli AA. 2021. Tracking the Swelling COVID-19 Vaccine Chatter on TikTok in Indonesia. Di dalam: Choong W, Lee PO, Lee SA, Meng NK, editor. *ISEAS Perspective*. Singapore(SG): Yusof Ishak Institute. hlm 1-12.
- Sholehah NA. 2018. Analisis Sentimen Menggunakan Naive Bayes pada Data Twitter Bahasa Indonesia. [skripsi]. Bogor(ID):Insitut Pertanian Bogor.
- Soumya S, Pramod KV. 2020. Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*. 6(4):300–305. doi: 10.1016/j.ict.2020.04.003
- Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. 2020. Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PLoS One*. doi:10.1371/journal.pone.0239441.
- Zheng Z, Cai Y, Li Y, Zheng Z, Cai Y, Li Y. 2015. Oversampling method for imbalanced classification. *Computing and Informatics*. 34:1017-1037.