

Analisis Sentimen Masyarakat Terhadap Penggunaan Aplikasi PeduliLindungi untuk Aktivitas Ruang Publik pada Media Sosial Twitter

1st Gilang Erlangga
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
gilangerlangga@student.telkomuni-
versity.ac.id

2nd Edi Sutoyo
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
edisutoyo@telkomuniversity.ac.id

3rd Hanif Fakhurroja
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
haniff@telkomuniversity.ac.id

Abstrak—Pemerintah Indonesia telah mengeluarkan kebijakan baru untuk membantu memantau penyebaran virus COVID-19 agar dapat ditangani secara efisien dengan melacak riwayat lokasi yang dikunjungi masyarakat dengan memerintahkan kepada mereka untuk melakukan check-in dan check-out di ruang publik menggunakan aplikasi PeduliLindungi. Twitter merupakan media yang digunakan sebagai sarana menyampaikan pendapat terkait isu yang hangat diperbincangkan. Beragam pendapat dikeluarkan oleh masyarakat baik bernada positif maupun negatif. Maka dari itu, dilakukan analisis sentimen terhadap penggunaan aplikasi PeduliLindungi. Analisis sentimen yang dilakukan menggunakan pendekatan *Machine Learning* dengan algoritma *Support Vector Machine* (SVM) sebagai metode untuk mengklasifikasikan setiap tweets nya ke dalam dua sentimen, yaitu positif dan negatif. Penelitian menghasilkan model terbaik yaitu menggunakan dengan metode pemisahan *train* dan *test* data yaitu metode holdout rasio 70:30, dan metode SMOTE untuk penyeimbangan datanya. Model tersebut menghasilkan nilai *accuracy*: 81.23%, *precision*: 83.58%, *recall*: 86.62%, *F-1 score*: 85.07%, dan *AUC*: 0.881. Model tersebut digunakan untuk klasifikasi penentuan sentimen dari setiap tweets sehingga dapat diketahui tren persebaran sentimen masyarakat terhadap penggunaan aplikasi PeduliLindungi untuk aktivitas ruang publik pada media sosial Twitter periode Oktober 2021 hingga Juli 2022.

Kata kunci— PeduliLindungi, Twitter, Analisis Sentimen, *Support Vector Machine* (SVM)

I. PENDAHULUAN

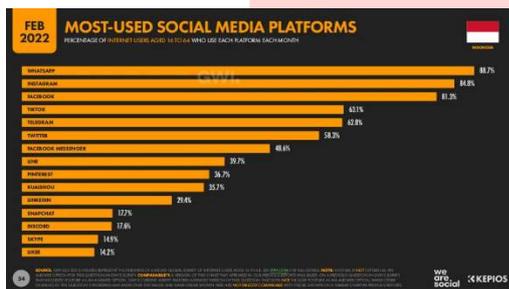
Saat ini, kehidupan di seluruh penjuru dunia telah mengalami perubahan yang sangat besar, mulai dari kebiasaan baru memakai masker, menjalani kegiatan pembelajaran dan bekerja secara jarak jauh, membatasi kegiatan di luar rumah, dan lain-lain. Perubahan tersebut mulai terjadi setelah

mewabahnya sebuah penyakit berbahaya yang mengancam kehidupan manusia, yaitu COVID-19. WHO menyatakan bahwa “COVID-19 merupakan penyakit yang disebabkan oleh virus corona baru yang disebut SARS-CoV-2 dan ditemukan pertama kali pada 31 Desember 2019 di Wuhan” [1]. Virus ini kemudian mulai masuk ke Indonesia pada tanggal 2 Maret 2020 melalui dua orang WNI yang sebelumnya melakukan kontak dengan warga Jepang, sebagaimana dikonfirmasi oleh Presiden Joko Widodo [2]. Kasus masyarakat terkonfirmasi positif serta meninggal akibat virus COVID-19 pun terus meningkat dan menyebar ke seluruh dunia hingga COVID-19 dinyatakan sebagai pandemi secara resmi pada bulan Maret 2020 [3]. Pada bulan Juli 2021, Indonesia mencatatkan bulan dengan kasus kematian terbanyak selama pandemi COVID-19 dengan jumlah kumulatif 32.061 kasus, sebagaimana disampaikan oleh Satuan Tugas Penanganan COVID-19 [4].

Dalam upaya penanggulangan pandemi COVID-19 yang terus meningkat, pemerintah mengeluarkan berbagai macam kebijakan, salah satunya adalah penerapan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Di samping penerapan PPKM, pemerintah mengeluarkan kebijakan yang mewajibkan masyarakat untuk melakukan screening menggunakan aplikasi PeduliLindungi saat beraktivitas di ruang publik, seperti supermarket dan hypermarket, mulai tanggal 14 September 2021 sebagaimana diatur dalam Inmendagri Nomor 39/2021 tentang PPKM Level 4, 3, dan 2 COVID-19 di Wilayah Jawa dan Bali [5]. Aktivitas tersebut dilakukan agar memudahkan pemerintah dalam melakukan pelacakan terkait kondisi kesehatan masyarakat yang berkunjung pada tempat umum.

Pada era Teknologi Informasi saat ini, manusia tidak dapat terlepas dari penggunaan internet. Salah satu aktivitas yang dilakukan dengan

memanfaatkan internet yaitu melakukan interaksi pada media sosial. Indonesia saat ini terdiri dari 277.7 juta penduduk, lalu 204.7 juta penduduk diantaranya menggunakan internet, dan 191.4 juta penduduk adalah pengguna media sosial yang aktif [6]. Media sosial merupakan media atau ruang digital yang memungkinkan pengguna untuk melakukan penyebaran konten dan komunikasi kepada pengguna yang lebih luas [7]. Dari beragam jenis platform media sosial yang digunakan di Indonesia, Twitter merupakan salah satu platform dengan jumlah pengguna aktif yang cukup banyak yaitu 58.3% dari total pengguna aktif media sosial [6].



GAMBAR 1 Survei media sosial di Indonesia

Twitter dapat dijadikan sebagai sarana untuk menyampaikan pendapat terkait dengan berbagai macam isu yang terjadi. Isu penggunaan aplikasi PeduliLindungi merupakan isu yang hangat diperbincangkan saat ini. Kebijakan yang dikeluarkan oleh pemerintah terkait penggunaan aplikasi PeduliLindungi, tentu saja mengeluarkan reaksi dan pendapat yang bermacam-macam dari masyarakat. Pada penelitian ini, akan dilakukan penggalian berbagai macam sentimen masyarakat dengan menganalisis komentar yang membahas atau menyikapi kebijakan serta penggunaan aplikasi PeduliLindungi untuk aktivitas ruang publik pada media sosial Twitter.

Analisis Sentimen yang dilakukan pada penelitian ini yaitu menggunakan pendekatan Machine Learning dengan metode klasifikasi. Algoritma klasifikasi yang digunakan pada penelitian ini yaitu Support Vector Machine (SVM) dengan kernel linear dan teknik SMOTE untuk mengatasi data yang tidak seimbang. Teknik klasifikasi yang dilakukan yaitu binary classification yang menghasilkan dua output dalam klasifikasi yaitu sentiment positif dan negatif. Hasil dari seluruh proses pengumpulan, pengolahan, dan analisis yang dilakukan pada penelitian ini berupa sajian informasi dalam bentuk visualisasi data terkait dengan sentimen masyarakat terhadap penggunaan aplikasi PeduliLindungi untuk aktivitas ruang publik pada media sosial Twitter.

II. KAJIAN TEORI

A. PeduliLindungi

PeduliLindungi merupakan aplikasi yang dikembangkan pada saat pandemi COVID-19. Aplikasi ini secara resmi dirilis oleh pemerintah berdasarkan Keputusan Menteri Komunikasi dan Informatika No. 177 Tahun 2020 tentang Penetapan Aplikasi PeduliLindungi Dalam Rangka Pelaksanaan Surveilans Kesehatan Penanganan COVID-19 [8].

Aplikasi ini dikembangkan dengan tujuan untuk pelacakan yang membutuhkan keterlibatan masyarakat dalam membagikan data lokasinya. Hasil pelacakan tersebut membantu memudahkan pemerintah dalam mengidentifikasi kondisi kesehatan masyarakat, lokasi terkait keramaian serta zonasi penyebaran COVID-19 [9].

B. Text Mining

Text mining didefinisikan sebagai teknik yang digunakan untuk menganalisis teks dalam jumlah yang besar agar menghasilkan sebuah pola, yang kemudian pola tersebut diekstrak menjadi suatu informasi yang berguna [10]. *Text mining* merupakan teknik yang terdiri dari penggabungan disiplin ilmu seperti *information retrieval*, *information extraction*, *data mining*, *machine learning*, *text analysis*, *classification*, dan *visualization* [11]. Informasi dan wawasan yang diperoleh dari berbagai macam sumber data, merupakan hal yang bernilai untuk membantu dalam pengambilan keputusan.

Text mining dapat digunakan dalam menganalisis media sosial. Karena media sosial merupakan media yang menghasilkan banyak sekali unstructured data terutama data tekstual. Salah satu teknik pengaplikasian *text mining* yang dapat digunakan pada media sosial adalah *text classification*. *Text classification* merupakan teknik pengelompokan data ke dalam sebuah kelas atau label.

C. Natural Language Processing (NLP)

Natural Language Processing adalah teknik komputasi yang digunakan untuk menganalisis dan mengidentifikasi teks berdasarkan seperangkat teori dan teknologi untuk membuat komputer bertindak seperti manusia dalam memproses dan memahami bahasa alami pada berbagai tugas dan aplikasi [12]. *Natural Language Processing* (NLP) menggabungkan teknik-teknik interdisiplin yang melibatkan konsep pada ilmu komputer, linguistik, logika dan psikologi [13].

NLP dapat digunakan untuk memanipulasi teks dengan mengambil nilai atau informasi penting di dalamnya. Sebagai contoh, adalah sistem pemrosesan teks otomatis. Sistem pemrosesan teks otomatis umumnya mengambil beberapa bentuk input teks dan mengubahnya menjadi output dari bentuk yang berbeda. Maka dari itu, perlu melewati berbagai macam tahapan untuk dapat mengekstraksi

atau memperoleh informasi penting dari data tekstual yang besar.

D. Text Pre-processing

Text Preprocessing merupakan proses paling penting pada text mining yang dilakukan untuk menyesuaikan format yang berbeda-beda dari setiap teks untuk membantu memudahkan pengolahan data pada pemrosesan selanjutnya [14]. Untuk menghasilkan model yang baik, maka diperlukan data yang berkualitas tinggi, maka dari itu text preprocessing merupakan langkah yang dilakukan dalam mempersiapkan data sebelum dilakukan pemodelan.

Pada penelitian [15] menyatakan terdapat beberapa tahapan pada text preprocessing yang efektif untuk dilakukan, diantaranya yaitu:

- *Convert text to lower case*: merupakan tahapan yang dilakukan untuk mengubah semua ukuran huruf menjadi huruf kecil.
- *Removal of @mention*: merupakan tahapan yang dilakukan untuk menghilangkan penyebutan username pada tweets.
- *Removal of URL*: merupakan tahapan yang dilakukan untuk menghilangkan tautan pada tweets.
- *Removal of punctuation*: merupakan tahapan yang dilakukan untuk menghapus tanda baca atau karakter non-alphanumeric pada tweets.
- *Removal of hashtag*: merupakan tahapan yang dilakukan untuk menghilangkan tagar pada tweets.
- *Removing whitespace*: merupakan tahapan yang dilakukan untuk menghapus baris baru pada tweets, karena baris baru tidak memberikan arti apapun pada teks.
- *Removal of the encoded text formats*: merupakan tahapan yang dilakukan untuk menghilangkan teks-teks yang dikodekan pada web, seperti: <, &, dan lain-lain.
- *Stop word removal*: merupakan tahapan yang dilakukan untuk menghilangkan kata-kata yang dianggap tidak penting.
- *Stemming*: merupakan tahapan yang dilakukan untuk mengubah bentuk kata menjadi kata dasarnya.
- *Spelling correction*: merupakan tahapan yang dilakukan untuk mengubah kata-kata yang tidak baku.

E. Synthetic Minority Oversampling Technique (SMOTE)

Imbalance Class atau ketidakseimbangan kelas merupakan masalah yang cukup sering dihadapi dalam *data mining* yang juga berlaku pada text mining. *Imbalance class* merupakan kondisi yang terjadi ketika salah satu kelas memiliki proporsi jumlah data yang sangat rendah dan sebaliknya

untuk kelas lainnya [16]. Dengan kata lain, *imbalance class* memiliki persebaran data yang tidak seimbang pada setiap kelasnya.

Synthetic Minority Oversampling Technique (SMOTE) merupakan teknik menambahkan data pada kelas minoritas secara sintesis yang berarti mengambil setiap sampel pada kelas minoritas dan memasukan contoh-contoh sintesis sepanjang segmen garis yang menghubungkan setiap tetangga terdekat pada kelas minoritas [17].

F. Term Frequency – Inverse Document Frequency (TF-IDF)

Algoritma Machine Learning, umumnya menerima masukan untuk dilakukan pemodelan berupa data numerik. Sedangkan, teks atau dokumen merupakan sebuah data tekstual. Maka dari itu, perlu melalui sebuah tahapan yang disebut dengan Vektorisasi. Vektorisasi merupakan proses mentransformasikan sebuah dokumen atau data tekstual menjadi sebuah vektor atau data numerik yang bertujuan agar dapat berkomunikasi dengan mesin untuk menyelesaikan masalah secara matematis [18]. Salah satu metode yang dapat digunakan dalam melakukan proses Vektorisasi, yaitu *Term Frequency – Inverse Document Frequency* (TF-IDF).

Term Frequency – Inverse Document Frequency merupakan metode pembobotan statistika yang digunakan untuk mengevaluasi seberapa penting kata pada sebuah dokumen dalam kumpulan dokumen. TF-IDF bekerja dengan cara menghitung nilai untuk setiap kata pada dokumen melalui proporsi terbalik dari frekuensi kata dalam dokumen tertentu [19]. Sebuah kata memiliki nilai TF IDF yang tinggi berarti memiliki hubungan yang kuat dengan dokumen yang memiliki kata tersebut.

G. Holdout Method

Dalam melakukan pembuatan model pada *machine learning*, terdapat suatu aktivitas untuk dilakukan validasi. Aktivitas tersebut dilakukan dengan memisahkan data menjadi dua bagian, yaitu *train set* atau data yang digunakan untuk melakukan pelatihan pada mesin dan *test set* atau data yang digunakan untuk menguji performa dari model yang dihasilkan. Dalam melakukan pembagian data tersebut, metode yang paling sering dan mudah untuk digunakan yaitu *holdout method*.

Holdout Method merupakan metode yang bekerja dengan hanya memisahkan atau membagi rasio dari *train set* dan *test set*, sehingga kelebihan dari metode ini yaitu tidak memerlukan waktu yang lama untuk melakukan validasi [20]. Rasio yang umum digunakan dalam melakukan pembagian untuk *train set* dan *test* yaitu 70:30, 75:25, dan 80:20.

H. K-Fold Cross Validation

Metode validasi yang juga dilakukan untuk melakukan pembagian menjadi *train set* dan *test set*

yaitu *K-Fold Cross Validation*. Metode ini dilakukan dengan melakukan pembagian secara acak untuk setiap data ke dalam *k* atau *fold* dengan ukuran yang sama [21].

I. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma supervised machine learning yang digagas pada tahun 1990-an dan umumnya dapat digunakan untuk pengenalan pola dalam mengklasifikasikan data ke dalam kelas [22]. SVM merupakan teknik yang bekerja untuk mendapatkan *hyperplane* yang paling mungkin untuk memisahkan dua kelas dengan mengukur margin *hyperplane* tersebut dan menentukan titik maksimumnya [23]. *Hyperplane* merupakan pemisah pada dimensi lebih dari tiga, sedangkan untuk tiga dimensi dinamakan *plane*, dan dalam dua dimensi, pemisah dinamakan garis.

J. Confusion Matrix

Salah satu metode pengukuran yang dapat digunakan dalam mengetahui performa dari sebuah model klasifikasi yaitu menggunakan confusion matrix. *Confusion Matrix* merupakan sebuah metrik $k \times k$, yang dimana *k* merupakan jumlah class yang terlibat dalam model klasifikasi [24]. Baris horisontal yang terdapat pada metrik ini merupakan nilai pada kelas prediksi. Sedangkan kolom vertikal pada metrik ini merupakan nilai pada kelas sebenarnya.

K. Receiver Operating Characteristic (ROC)

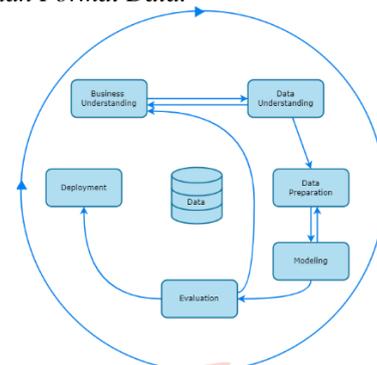
Metode pengukuran lain yang komprehensif digunakan untuk mengevaluasi model klasifikasi yaitu kurva ROC. *Receiver Operating Characteristic* merupakan kurva yang menggambarkan hasil klasifikasi dari kelas *true positive* dan kelas *false positive* [25].

III. METODE

Pada penelitian ini, sistematika penyelesaian masalah yang dilakukan adalah dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* atau yang lebih dikenal dengan akronimnya, yaitu CRISP-DM [26].

Berdasarkan GAMBAR 2 di bawah, pada metode CRISP-DM ini terbagi menjadi enam fase yang dapat bergerak maju mundur. Fase yang pertama, yaitu *Business Understanding*. Fase ini merupakan fase awal yang dilakukan untuk mengidentifikasi permasalahan dan menghasilkan sebuah rencana untuk mencapai tujuan pemecahan masalah. Fase yang kedua, yaitu *Data Understanding*. Setelah menemukan sebuah rencana yang bertujuan untuk memecahkan masalah, tahap selanjutnya adalah memahami data-data apa saja yang diperlukan dalam melakukan analisis sentimen. Setelah memahami data-data apa saja yang diperlukan, maka fase selanjutnya adalah menyiapkan data-data yang diperlukan atau *Data Preparation*. Pada fase ini, terdiri dari beberapa

proses, yaitu *Select Data*, *Clean Data*, *Integrate Data*, dan *Format Data*.



GAMBAR 2 CRISP-DM

Setelah data sudah siap untuk diolah, maka selanjutnya beralih pada fase yang keempat, yaitu *Modelling*. Fase ini merupakan fase yang dilakukan untuk menemukan pola dari dataset menggunakan algoritma klasifikasi dan algoritma pembobotan teks atau dokumen. Setelah model berhasil dibuat, maka fase yang selanjutnya yaitu fase *Evaluation*, yang bertujuan untuk mengetahui performa dari model klasifikasi yang telah dibuat. Apabila metode klasifikasi yang dilakukan dalam melakukan *Modelling* telah menjawab permasalahan yang telah diidentifikasi pada tahap awal dan proses *Evaluation* sudah berjalan dengan baik, maka proses selanjutnya adalah *Deployment*, yaitu menyajikan informasi yang telah diperoleh ke dalam sebuah bentuk yang mudah dimengerti oleh banyak orang. Namun, jika model yang dihasilkan belum sesuai, maka perlu dilakukan perbaikan dengan kembali meninjau ulang mulai dari fase *Business Understanding* hingga menemukan hasil yang sesuai.

IV. HASIL DAN PEMBAHASAN

A. Business Understanding

Objek pada penelitian ini yaitu penggunaan aplikasi PeduliLindungi pada platform Twitter. Penelitian ini memanfaatkan data-data tweets dari masyarakat pengguna Twitter yang berkaitan dengan penggunaan aplikasi PeduliLindungi dari rentang waktu 8 Oktober 2021 sampai dengan 13 Juli 2022. *Tweets* tersebut kemudian dilakukan pengelompokan berdasarkan sentimen nya yaitu Positif dan Negatif dengan menggunakan pendekatan *Machine Learning*.

Pada implementasinya, dibutuhkan sebuah model yang dapat digunakan untuk menemukan pola dari kumpulan data yang dimiliki. Model *Machine Learning* tersebut dirancang dengan menggunakan algoritma *Support Vector Machine (SVM)* dengan *kernel linear*. Dengan memanfaatkan penggunaan model *Machine Learning* tersebut, maka diharapkan dapat dilakukan analisis sentimen masyarakat terhadap penggunaan aplikasi PeduliLindungi yang bertujuan untuk mengetahui hasil sentimen

masyarakat terkait dengan penggunaan aplikasi PeduliLindungi untuk aktivitas ruang publik pada media sosial Twitter.

B. Data Understanding

Data yang digunakan pada penelitian ini yaitu data cuitan (*tweets*) masyarakat yang berkaitan dengan penggunaan aplikasi PeduliLindungi. Data diambil dengan memanfaatkan API menggunakan *library* dari Python yaitu Tweepy. Data diambil dengan beragam kata kunci yang berkaitan dengan objek penelitian. Jumlah keseluruhan data yang telah dikumpulkan yaitu sebanyak 8.004 data. Data tersebut merupakan data terpilih yang tidak mengandung duplikat dan juga tidak mengandung *retweet* post. Data-data tersebut masih bersifat *raw data* sehingga pada fase selanjutnya akan dilakukan persiapan pada data untuk menghasilkan kualitas data yang baik. Atribut yang diambil yaitu atribut: *date*, *username*, dan *tweets*.

Dataset sebanyak 8.004 data tersebut kemudian dipecah menjadi data yang akan dilakukan pemodelan dan data yang akan dilakukan prediksi terhadap kelas sentimennya. Sebanyak 5.021 data digunakan untuk pemodelan, sedangkan untuk dilakukan prediksi terdiri sebanyak 2.983 data. Model *Machine Learning* yang digunakan yaitu *Supervised Machine Learning* sehingga membutuhkan sebuah atribut baru bernama label yang menjadi penentu sentimen pada setiap *tweets* dalam dataset yang akan dibuat pada tahap selanjutnya.

C. Data Preparation

Data yang digunakan dalam pembuatan model *Machine Learning* mempengaruhi kualitas model yang dihasilkan. Agar dapat menghasilkan model yang baik, maka perlu dilakukan persiapan pada *dataset*. Proses *Data Cleaning* yang efektif untuk dilakukan dalam menangani kasus data tekstual diantaranya yaitu: membuat atribut label, menyesuaikan ukuran huruf, mengilangkan baris baru, *username*, URL, karakter special pada HTML, tagar, tanda baca, emoji, angka, kata-kata tidak penting, mengoreksi kata ke dalam bentuk kata baku serta mengubah bentuk kata menjadi kata dasar.

D. Modelling

Setelah dilakukan berbagai macam proses pada fase *Data Preparation*, selanjutnya data akan diolah lebih lanjut dan dilakukan pemodelan sebagai alat prediksi sentimen. Proses-proses pada fase *modeling* ini yaitu: *data splitting*, *class balancing*, dan *data training*.

Pada proses *data splitting*, akan dilakukan dua metode yaitu metode *Holdout* dan *K-Fold Cross Validation*. Proses dilakukan dengan memecah dataset menjadi dua bagian, yaitu *train set* dan *test set*. *Train set* merupakan kumpulan data yang digunakan untuk melatih model, sedangkan *test set* merupakan kumpulan data yang digunakan untuk

menguji performa dari model yang telah dilatih. Pada metode *Holdout* dataset dipecah dengan rasio 70:30, 75:30, dan 80:20 secara *stratified*.

Sedangkan pada metode *K-Fold Cross Validation*, setiap data pada keseluruhan dataset mengalami fase training dan testing dengan melakukan iterasi sebanyak 10 kali. Iterasi tersebut terdiri atas 10 fold, yang dimana pada iterasi ke 1, fold 1 yang digunakan sebagai data testing, sedangkan fold 2 hingga fold 10 merupakan data training. Lalu pada iterasi ke 2, fold 2 yang digunakan sebagai data testing, sedangkan fold 1 lalu fold 3 hingga fold 10 merupakan data training. Proses tersebut dilakukan hingga iterasi ke 10.

Pada proses *class balancing*, dilakukan proses penyeimbangan pada data *training*, karena dataset pada penelitian ini memiliki karakteristik persebaran label yang tidak seimbang. Metode penyeimbangan data yang digunakan pada penelitian ini yaitu SMOTE.

TABEL 1 persebaran data sebelum *Class Balancing*

Rasio	Sebelum SMOTE		
	<i>Negative</i>	<i>Positive</i>	Jumlah
70:30	1.383	2.459	3.842
75:25	1.482	2.634	4.116
80:20	1.581	2.810	4.391

Pada TABEL 1 di atas, merupakan contoh persebaran data sebelum dilakukan *class balancing* atau SMOTE.

TABEL 2 persebaran data setelah *Class Balancing*

Rasio	Sesudah SMOTE		
	<i>Negative</i>	<i>Positive</i>	Jumlah
70:30	2.459	2.459	4.918
75:25	2.634	2.634	5.268
80:20	2.810	2.810	5.620

Pada TABEL 2 di atas, telah dilakukan penyeimbangan data dengan menggunakan SMOTE berdasarkan masing-masing rasio. Data minoritas dilakukan penambahan data secara sintesis, sehingga persebaran data minoritas dan mayoritas menjadi seimbang.

Selanjutnya, data-data yang sudah melalui berbagai macam proses selanjutnya dilakukan training data, untuk pembangunan sebuah model prediksi. Algoritma yang digunakan dalam pembangunan model tersebut yaitu *Support Vector Machine* (SVM) dengan kernel linear. Namun dalam pembagian data training dan testing, pada penelitian ini dilakukan dua metode, yaitu metode *holdout* dan *K-fold Cross Validation*. Serta dalam menghadapi dataset yang tidak seimbang, maka dilakukan percobaan untuk mengabaikan data tersebut imbalance dan dilakukan pendekatan SMOTE.

Setelah dilakukan pemodelan dengan berbagai macam metode, maka selanjutnya akan dilakukan evaluasi pada model untuk mengetahui performa

klasifikasi yang dihasilkan dengan melakukan pengujian terhadap data testing. Pada penelitian ini, metode evaluasi yang digunakan untuk mengukur performa klasifikasi yaitu dengan menggunakan confusion matrix, dan ROC curve dan metrik penilaian klasifikasi.

Metriks pengukuran yang dapat digunakan dalam mengukur performa model klasifikasi yang telah dihasilkan yaitu dengan menggunakan accuracy, recall, precision, F-1 score, dan AUC yang dapat ditentukan setelah diketahui hasil dari nilai confusion matrix sebelumnya. Berikut pada Tabel V. 3 merupakan hasil performa model yang dihasilkan dari berbagai macam metode yang digunakan.

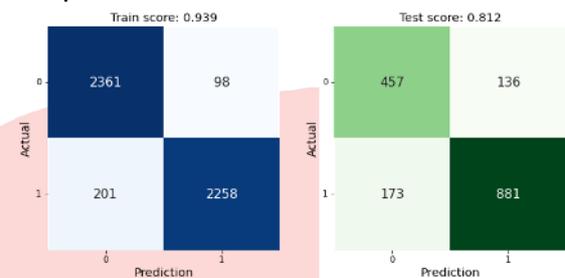
TABEL 3 Metriks Penilaian Klasifikasi

Metode	Accuracy	Recall	Precision	F-1 score	AUC
Hold-out 70:30	0.810	0.879	0.833	0.855	0.885
Hold-out 75:25	0.811	0.879	0.834	0.856	0.884
Hold-out 80:20	0.808	0.874	0.834	0.854	0.890
K-Fold	0.888	0.859	0.819	0.838	0.971
Hold-out 70:30 SM-OTE	0.812	0.835	0.866	0.850	0.881
Hold-out 75:25 SM-OTE	0.800	0.820	0.861	0.840	0.877
Hold-out 80:20 SM-OTE	0.806	0.827	0.863	0.845	0.886
K-Fold SM-OTE	0.783	0.806	0.847	0.825	0.972

Berdasarkan TABEL 3 di atas, maka diperoleh model dengan nilai accuracy dan precision tertinggi yaitu model dengan menggunakan metode holdout rasio 70:30 dan SMOTE dengan nilai accuracy sebesar 81.23% dan nilai precision sebesar 86.62%. Selanjutnya, untuk nilai recall tertinggi dihasilkan oleh model yang dibiarkan imbalance dengan metode holdout rasio 70:30, untuk nilai F-1 score tertinggi dihasilkan oleh model yang dibiarkan imbalance dengan metode holdout rasio 75:30, dan

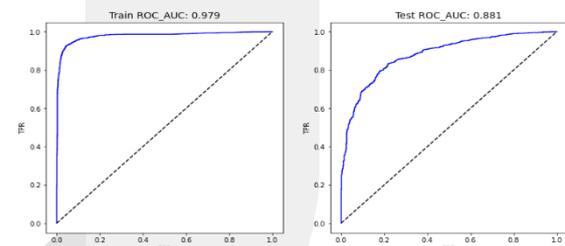
nilai AUC tertinggi dihasilkan oleh model dengan menggunakan metode K-fold cross validation dan SMOTE dengan nilai AUC sebesar 0.972.

Model terbaik yang digunakan untuk melakukan klasifikasi sentimen masyarakat terhadap penggunaan aplikasi PeduliLindungi pada aktivitas ruang publik yaitu model dengan menggunakan algoritma Support Vector Machine (SVM) linear kernel dengan metode Holdout skala 70:30 dan metode SMOTE.



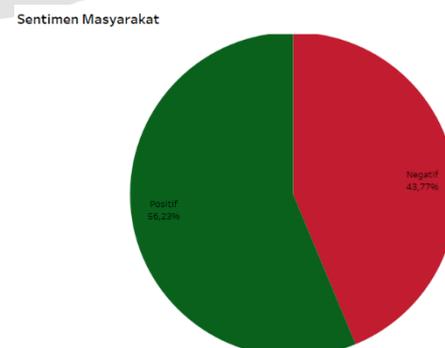
GAMBAR 3 Confusion Matrix Holdout 70:30 SMOTE

Pada GAMBAR 3 di atas, merupakan confusion matrix yang dihasilkan dari model dengan metode holdout skala 70:30 dan SMOTE. Model tersebut pada fase training mendapatkan nilai TN sebanyak 2361, FP sebanyak 98, FN sebanyak 201, dan TP sebanyak 2258. Sementara pada fase testing mendapatkan nilai TN sebanyak 408, FP sebanyak 185, FN sebanyak 127, dan TP sebanyak 927.



GAMBAR 4 ROC curve Holdout skala 70:30 SMOTE

Pada GAMBAR 4 di atas, merupakan ROC curve yang dihasilkan dari model dengan metode Holdout skala 70:30 SMOTE. Model tersebut mendapatkan nilai AUC sebesar 0.979 pada fase training dan sebesar 0.881 pada fase testing.



GAMBAR 5 Pie chart sentimen masyarakat

Sedangkan untuk sentimen bernada negatif pada GAMBAR 8, persebaran kata dengan frekuensi terbanyak juga yaitu pada kata ‘pedulilindungi’ dan ‘aplikasi’ karena seluruh tweets pada sentimen negatif berbicara tentang aplikasi PeduliLindungi. Persebaran kata terbanyak lainnya yaitu pada kata ‘enggak’. Sebagian besar tweets yang mengandung kata ‘enggak’ atau ‘tidak’ membicarakan aplikasi PeduliLindungi yang tidak dapat digunakan, sertifikat vaksin tidak muncul, dan beberapa fungsi pada aplikasi yang tidak dapat digunakan.

Lalu, ditemukan kata-kata yang membicarakan teknis aplikasi seperti ‘error’, ‘update’, dan ‘scan’. Kata ‘error’ merupakan kata yang merujuk pada aplikasi yang tidak berfungsi dengan semestinya atau mengalami gangguan. Lalu, sebagian besar tweets yang mengandung kata ‘update’ berisikan bahwa aplikasi mengalami penurunan performa, rentan error, dan mengeluhkan fungsi-fungsi vital yang dihilangkan setelah dilakukan pembaruan pada aplikasi PeduliLindungi, sehingga pengguna banyak mengeluhkan kondisi setelah dilakukan pembaruan pada aplikasi.

Lalu untuk kata ‘scan’, sebagian besar dapat dihubungkan dengan kata sebelumnya yaitu ‘enggak’. Dalam beberapa tweets mengandung kata yang berisi ‘enggak bisa scan’, yang berarti fungsi utama dalam aplikasi ini yaitu melakukan scan QR code untuk melakukan aktivitas *check-in* dan *check-out* terkendala pada penggunaan aplikasi PeduliLindungi ini.

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan untuk hasil akurasi tertinggi yang dihasilkan dan dipilih sebagai model terbaik untuk melakukan klasifikasi sentimen yaitu dengan nilai *accuracy* sebesar 81.23%, *precision* sebesar 83.58%, *recall* sebesar 86.62%, *F-1 score* sebesar 85.07%, dan nilai AUC sebesar 0.881. Hasil akurasi tersebut dihasilkan oleh model dengan penggunaan algoritma *Support Vector Machine* (SVM) menggunakan metode *Holdout* rasio 70:30 untuk pembagian datanya dan metode SMOTE untuk penyeimbangan datanya.

Berdasarkan data yang dianalisa sebanyak 8.004 dari 5.695 pengguna dan rentang waktu mulai dari 8 Oktober 2021 sampai dengan 13 Juli 2022, hasil sentimen masyarakat terkait penggunaan aplikasi PeduliLindungi untuk aktivitas ruang publik pada media sosial Twitter yaitu sebanyak 56.23% bersentimen positif dan 43.77% bersentimen negatif. Dengan persebaran yang naik turun untuk masing-masing sentimennya pada setiap bulan.

REFERENSI

[1] World Health Organization (WHO), “Coronavirus disease (COVID-19),” 2021. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19)

[answers-hub/q-a-detail/coronavirus-disease-covid-19](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19) (accessed Jan. 25, 2022).

- [2] T. T. S. Berty, “Kasus Pertama Virus Corona di Indonesia Jadi Sorotan Dunia,” Mar. 02, 2020.
- [3] World Health Organization (WHO), “WHO Director-General’s opening remarks at the media briefing on COVID-19,” 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed Jan. 25, 2022).
- [4] CNN Indonesia, “Juli 2021, Bulan dengan Kasus Kematian Covid-19 Terbanyak,” Jul. 29, 2021.
- [5] H. D. Situmorang, “Masuk Supermarket Wajib Gunakan Aplikasi PeduliLindungi Mulai 14 September,” Sep. 07, 2021.
- [6] We Are Social, “Digital 2022: Indonesia,” 2022. <https://datareportal.com/reports/digital-2022-indonesia>.
- [7] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, “Advances in Social Media Research: Past, Present and Future,” *Inf. Syst. Front.*, vol. 20, no. 3, pp. 531–558, 2018, doi: 10.1007/s10796-017-9810-y.
- [8] R. Fauzan, “Aplikasi PeduliLindungi Resmi Tersedia di Playstore dan Appstore,” Apr. 14, 2020.
- [9] PeduliLindungi, “PeduliLindungi,” 2020. <https://www.pedulilindungi.id/#tentang>.
- [10] F. Sebastiani, “Machine Learning in Automated Text Categorization,” *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, 2002, doi: 10.1145/505282.505283.
- [11] A. Tan, “Text Mining : The state of the art and the challenges Concept-based,” *Proc. PAKDD 1999 Work.*, no. January, pp. 65–70, 2011, [Online]. Available: <http://www.mendeley.com/research/text-mining-state-art-challenges-3/>.
- [12] E. D. Liddy, “Natural language processing,” *Encycl. Libr. Inf. Sci.*, pp. 131–136, 2001, doi: 10.1016/0004-3702(82)90032-7.
- [13] A. K. Josh, “Natural language processing,” *Science (80-.)*, vol. 253, no. 5025, pp. 1242–1249, 1991, doi: 10.1126/science.253.5025.1242.

- [14] M. N. Dr. S. Vijayarani, Ms. J. Ilamathi, "Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. October 2014, pp. 7–16, 2015.
- [15] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective text data preprocessing technique for sentiment analysis in social media data," *Proc. 2019 11th Int. Conf. Knowl. Syst. Eng. KSE 2019*, pp. 1–8, 2019, doi: 10.1109/KSE.2019.8919368.
- [16] M. Kuhn and K. Johnson, *Applied Predictive Modeling with Applications in R*, vol. 26. 2013.
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, no. Sept. 28, pp. 321–357, 2002, [Online]. Available: <https://arxiv.org/pdf/1106.1813.pdf> <http://www.snopes.com/horrors/insects/telamonias.asp>.
- [18] A. K. Singh and M. Shashi, "Vectorization of text documents for identifying unifiable news articles," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 305–310, 2019, doi: 10.14569/ijacsa.2019.0100742.
- [19] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proc. first Instr. Conf. Mach. Learn.*, vol. 242, no. 1, pp. 29–48, 2003.
- [20] M. A. A. Halim, M. T. A. Rahman, N. A. Rahim, A. Rahman, A. F. A. Hamid, and N. A. M. Amin, "Analysis on current flow style for vehicle alternator fault prediction," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 670, no. 1, 2019, doi: 10.1088/1757-899X/670/1/012042.
- [21] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer Texts in Statistics, 2013.
- [22] A. Pradhan, "SUPPORT VECTOR MACHINE-A Survey," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, pp. 82–85, 2012.
- [23] A. Turnip, M. F. Amri, H. Fakrurroja, A. I. Simbolon, M. A. Suhendra, and D. E. Kusumandari, "Deception detection of EEG-P300 component classified by SVM method," *ACM Int. Conf. Proceeding Ser.*, no. June 2019, pp. 299–303, 2017, doi: 10.1145/3056662.3056709.
- [24] A. D. Forbes, "Classification-algorithm evaluation: Five performance measures based on confusion matrices," *J. Clin. Monit.*, vol. 11, no. 3, pp. 189–206, 1995, doi: 10.1007/BF01617722.
- [25] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," *AAAI Work. - Tech. Rep.*, vol. WS-06-06, pp. 24–29, 2006, doi: 10.1007/11941439_114.
- [26] P. Chapman *et al.*, "Crisp-Dm 1.0," *Crisp Consort.*, p. 76, 2000.