Jurnal Teknik Informatika (JUTIF)
Vol. 3, No. 4, Agustus 2022, hlm. 807-814

DOI: https://doi.org/10.20884/1.jutif.2022.3.4.292
p-ISSN: 2723-3863
e-ISSN: 2723-3871

TWITTER SENTIMENT ANALYSIS PEDULILINDUNGI APPLICATION USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE

Indra Yunanto*[1], Sri Yulianto[2]

[1,2]Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia
Email: [1]672018019@student.uksw.edu, [2]sri.yulianto@uksw.edu

(Naskah masuk: 25 April 2022, Revisi : 29 April 2022, diterbitkan: 20 Agustus 2022)

Abstract

The PeduliLindungi application is an application launched by the government during the COVID-19 pandemic, with the aim of helping government agencies carry out digital tracking to monitor the public, as an effort to prevent the spread of the Corona virus. Many people express their opinions on the PeduliLindung application on social media, one of which is through Twitter. To improve the performance of the application, of course, need input or complaints from users, opinions from the public on Twitter about the PeduliLindungi application can be input to improve or improve the performance of the application. Sentiment analysis is carried out to see how the public's sentiment towards the PeduliLindung application is, and these sentiments will be categorized into positive sentiment and negative sentiment, this sentiment can later be used as evaluation material for application development. This study aims to see and compare the accuracy of two classification methods, Naïve Bayes and Support Vector Machine in the classification process of sentiment analysis. The data used are 4636 tweets with the keyword " PeduliLindungi". The data obtained then goes to the pre-processing stage before going to the classification stage. The results obtained after classifying using the Naïve Bayes method and the Support Vector Machine show that the Support Vector Machine method has a higher accuracy of 91%, while the Naïve Bayes method has an accuracy of 90%.

Keywords: *naïve bayes, pedulilindungi, sentiment analysis, support vector machine.*

ANALISIS SENTIMEN TWITTER APLIKASI PEDULILINDUNGI MENGGUNAKAN *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE*

Abstrak

Aplikasi PeduliLindungi merupakan aplikasi yang diluncurkan oleh pemerintah pada masa pandemi COVID-19, dengan tujuan membantu instansi pemerintah melakukan pelacakan digital untuk memantau masyarakat, sebagai upaya pencegahan penyebaran virus Corona. Banyak masyarakat menyampaikan opini mereka terhadap aplikasi PeduliLindungi di media sosial, salah satunya melalui *Twitter*. Untuk meningkatkan performa dari aplikasi tentu butuh masukkan atau keluhan dari pengguna, opini dari masyarakat pada *Twitter* terhadap aplikasi PeduliLindungi bisa menjadi masukkan untuk memperbaiki ataupun meningkatkan performa dari aplikasi. Sentimen analisis dilakukan untuk melihat bagaimana sentimen masyarakat terhadap aplikasi PeduliLindungi, dan sentimen tersebut akan dikategorikan menjadi sentimen positif dan sentimen negative, dari sentimen tersebut nantinya bisa menjadi bahan evaluasi untuk pengembangan aplikasi. Penilitian ini bertujuan untuk melihat dan membandingkan akurasi dari dua metode klasifikasi yaitu *Naïve Bayes* dan *Support Vector Machine* dalam proses klasifikasi analisis sentimen. Data yang digunakan adalah sebanyak 4636 *tweet* dengan kata kunci “PeduliLindungi”. Data yang didapat kemudian masuk ke tahap *pre-processing* sebelum nantinya masuk ke tahap klasifikasi. Hasil yang didapat setelah melakukan klasifikasi menggunakan metode *Naïve Bayes* dan *Support Vector Machine* menunjukkan bahwa metode *Support Vector Machine* memiliki akurasi yang lebih tinggi yaitu 91%, sedangkan metode *Naïve Bayes* memiliki akurasi sebesar 90%.

Kata kunci: *analisis sentimen, naïve bayes, pedulilindungi, support vector machine.*

1. PENDAHULUAN

Teknologi dan informasi berkembang semakin pesat, termasuk salah satunya adalah internet. Internet memberikan fasilitas yang dapat memudahkan masyarakat dalam memperoleh informasi dari berbagai penjuru dunia. Di Indonesia sendiri pengguna internet sangat banyak dan terus meningkat. Menurut laporan bertajuk “Digital 2021”

yang dirilis oleh layanan manajemen konten *HootSuite* dan agensi pemasaran media *sosial We Are Social*, pengguna internet di Indonesia pada awal 2021 ini mencapai 202,6 juta jiwa. Aktivitas berinternet yang paling disukai oleh pengguna internet di Indonesia adalah bermedia sosial. Merujuk dari hasil laporan yang sama, saat ini ada 170 juta jiwa orang Indonesia yang merupakan pengguna aktif media sosial. Menurut laporan yang dirilis oleh *HootSuite* dan *We Are Social*, *Twitter* berada di peringkat lima teratas media sosial yang digunakan di Indonesia. *Twitter* merupakan sebuah media sosial yang memungkinkan penggunanya untuk berinteraksi secara personal ataupun terbuka[1]. Salah satu topik yang diperbincangkan oleh masyarakat Indonesia pada masa pandemi adalah mengenai aplikasi PeduliLindungi.

Aplikasi PeduliLindungi merupakan aplikasi yang diluncurkan oleh pemerintah pada masa pandemi Covid-19. Aplikasi ini dikembangkan untuk membantu instansi pemerintah melakukan pelacakan digital untuk memantau masyarakat sebagai upaya untuk mencegah penyebaran virus Corona[2]. Sebagian besar masyarakat menyuarakan opini mereka mengenai aplikasi PeduliLindungi di platform media sosial *Twitter* sehingga sering menjadi *trending topic* pada platform tersebut. Opini tersebut merupakan data textual yang dapat dianalisa dan dimanfaatkan untuk mendapatkan informasi guna melihat sentimen masyarakat terhadap aplikasi PeduliLindungi. Untuk melihat bagaimana sentimen masyarakat terhadap aplikasi PeduliLindungi dapat dilakukan sentiment analysis, dan sentimen tersebut akan dikategorikan menjadi sentimen positif dan sentimen negatif sehingga dari hasil sentimen yang didapat bisa dimanfaatkan untuk evaluasi aplikasi PeduliLindungi.

Beberapa penelitian mengenai analisis sentimen di media sosial sudah pernah dilakukan sebelumnya. Salah satunya adalah penelitian yang dilakukan oleh Sarika Afrizal, dkk mengenai analisis sentimen warga Jakarta terhadap kehadiran MRT dengan metode klasifikasi *Naïve Bayes*. Penelitian tersebut berguna untuk melihat opini warga Jakarta terhadap MRT untuk nantinya menjadi bahan evaluasi bagi penyedia layanan. Hasil Akurasi yang didapat dengan metode *Naïve Bayes* sebesar 75% [3].

Penelitian yang dilakukan oleh Samsir, dkk dengan judul "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode *Naïve Bayes*" pada tahun 2021. Pada masa pendemi semua pendidikan menggunakan sistem daring dalam proses belajar mengajar. Banyak kontroversi dalam masyarakat yang terjadi akibat perubahan sistem belajar mengajar yang sebelumnya tatap muka menjadi daring. Penelitian ini bertujuan menganalisis opini publik terhadap pembelajaran daring pada masa pandemi di Indonesia pada awal November 2020. Penelitian dilakukan dengan mengambil opini masyarakat pada *Twitter* mengenai pembelajaran daring, dan didapatkan data sebanyak 12,906 twit yang akan dianalisis menggunakan algoritma *Naïve Bayes*. Pada penelitian ini diperoleh bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral pada periode tersebut, serta memiliki nilai akurasi dari *Naïve Bayes* sebesar 97,15% [4].

Penelitian yang dilakukan oleh Winda Yulita, dkk dengan judul "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma *Naïve Bayes Classifier*". Setelah pandemi berjalan cukup lama di Indonesia, akhirnya pada tanggal 9 November 2020 vaksin pertama diumumkan kepada masyarakat, dengan kefektifan vaksin lebih dari 90%. Masyarakat banyak menyampaikan opini mereka mengenai vaksin di media sosial, salah satunya adalah *Twitter*. Tujuan dari penelitian ini adalah untuk menganalisis pendapat masyarakat Indonesia mengenai vaksinasi COVID-19 di media sosial *Twitter*. Analisis dilakukan terhadap data 3780 *tweet* yang berkaitan vaksinasi dengan menggunakan algoritma *Naïve Bayes Classifier*. Berdasarkan hasil analisis, dapat dilihat bahwa sebagian besar *tweet* memiliki sikap positif (60,3%), sementara jumlah tweet yang netral (34,4%) melebihi jumlah tweet yang menentang (5,4%). Nilai akurasi yang dihasilkan dari algoritma *Naïve Bayes Classifier* sebesar 0,93 (93%) [5].

Penelitian yang dilakukan oleh Brian Laurensz dan Eko Sediyono dengan judul "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19". Pandemi melada Indonesia cukup lama, pemerintah mengambil Tindakan untuk melakukan vaksinasi agar menekan angka penyebaran Covid-19 di Indonesa. Tanggapan masyarakat mengenai vaksinasi sangat beragam, sebagian masyarakat menyampaikan pendapat mereka di sosial media *Twitter*. Pada penelitian ini metode yang digunakan adalah metode SVM dan Naïve Bayes. Penelitian ini bertujuan untuk mengetahui opini masyarakat di media sosial Twitter terhadap tindakan vaksinasi. Data yang digunakan dalam penelitian ini sebanyak 845 tweet, dengan menggunakan dua kata kunci, yaitu "vaksinmerahputih" dan "vaksinsinovac". Tahap yang dilakukan sebelum proses klasifikasi adalah *pre-processing* (*Transform Cases*, *Tokenize*, *Filter Token* (*By Length*), *Stopword Removal*, *Stemming*, *Generate n-Grams*), pembobotan tf-idf. Data kemudian dibagi menjadi 253 data latih dan 592 data uji. Setelah dilakukan klasifikasi metode *Naïve Bayes* mempunyai rata-rata tingkat akurasi lebih besar dengan persentase sebesar 85,59%, sedangkan metode SVM 84,41% [6].

Penelitian yang dilakuka oleh Fajar Sodik Pamungkas dan Iqbal Kharisudin dengan judul "Analisis Sentimen dengan SVM, *NAIVE BAYES* dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial

Twitter". Pandemi Covid-19 sangat berdampak diberbagai sektor kehidupan masyarakat, keadaan yang memaksa masyarakat untuk melaksanaan physical distancing merubah pola hidup masyarakat. Hal tersebut membuat berbagai pendapat atau tanggapan masyarakat terhadap pandemi Covid-19 yang dituangkan dalam media sosial. Pada penelitian ini dilakukan analisis sentimen tanggapan masyarakat Indonesia terhadap pandemi Covid-19 pada media sosial *Twitter* menggunakan algoritma *Support Vector Machine* (SVM), *Naive Bayes*, dan *K-Nearest Neighbor* , yang kemudian ketiga algoritma tersebut dibandingkan mana yang paling baik untuk mengklasifikasikan data tanggapan. Hasil akurasi dengan menggunakan evaluasi model 10-Fold Cross Validation , dapat disimpulkan bahwa algoritma SVM memiliki akurasi yang lebih tinggi daripada *Naive Bayes* dan KNN dengan rata - rata akurasinya sebesar 90,01% pada SVM dengan kernel linear, 79,20% pada *Naive Bayes* dengan jumlah *laplace* adalah 1, dan 62,10% pada KNN dengan jumlah K adalah 20 dan menggunakan kernel *optimal*[7].

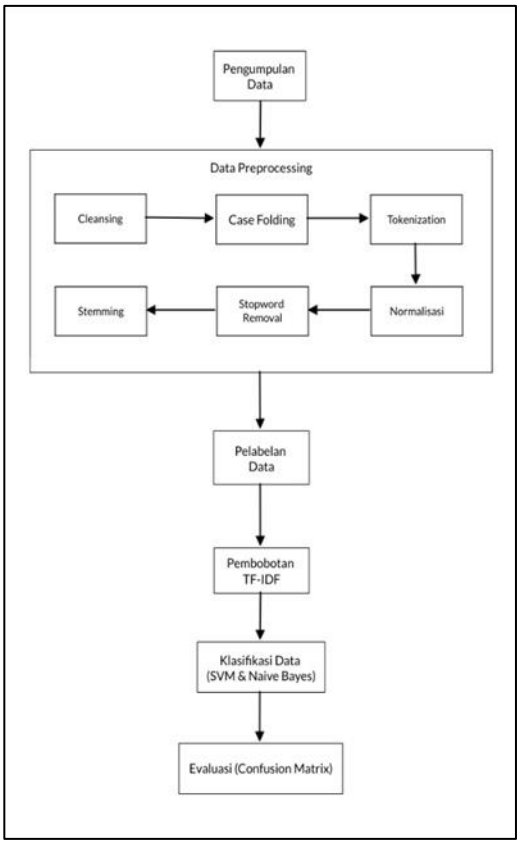
Penelitian yang sudah dilakukan sebelumnya bisa membuktikan bahwa analisis sentiment dapat dilakukan dengan menggunakan *machine learning*, salah satu metodenya adalah klasifikasi. Berdasarkan hal tersebut dan didukung dengan latar belakang masalah yang ada, penulis ingin meneliti tentang analisis sentimen *Twitter* yang berfokus pada aplikasi PeduliLindungi, dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* untuk melihat perbandingan akurasi dari dua metode tersebut.

2. METODE PENELITIAN

Text Mining atau yang biasa juga disebut *text data mining* adalah proses mengekstraksi pola atau pengetahuan yang menarik dari dokumen teks yang tidak terstruktur. Tugas yang dimiliki oleh text mining lebih kompleks dibanding data mining karena melibatkan data penanganan data teks yang inheren, tidak terstruktur, dan fuzzy. Text mining merupakan bidang multidisiplin, yang melibatkan pencarian informasi, analisis teks, ekstraksi informasi, pengelompokan, kategorisasi, visualisasi, database teknologi, *Machine learning*, dan data mining . *Text mining* bisa menyelesaikan permasalahan seperti pemrosesan, pengorganisasian atau pengelompokkan dan menganalisa teks yang tidak terstruktur dalam jumlah besar[8].

Analisis sentimen adalah cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence*. Analisis sentimen bertujuan untuk memahami, mengekstrak, dan mengloah data yang berbentuk teks secara otomatis hingga menghasilkan informasi yang berguna[9]. Analisis snetimen dalam penelitian ini adalah untuk mengklasifikasi sebuah data tekstual ke dalam dua kelas, yaitu negatif dan positif.

Metode penelitian yang dilakukan akan disajikan dalam bentuk gambar 1.



Gambar 1. Tahap penelitian

Tahap pertama dalam penelitian ini adalah pengumpulan data. Data yang diambil merupakan tweet dari *Twitter* yang mengandung kata "PeduliLindungi". Proses pengambilan data menggunakan tools bernama *Twint* yang ada di Python. *Twint* adalah tools scrapping tweets yang terdapat di Python yang memungkinkan untuk scrapping tweets dari profil Twitter tanpa menggunakan API *Twitter*. Pada tahap ini, tweet yang diambil adalah tweet dari tanggal 2021-09-28 hingga 2021-10-07. Setelah penarikkan data tweet dari Twitter kemudian data tersebut disimpan dalam format .csv.

Setelah melakukan pengumpulan data, tahap berikutnya adalah data preprocessing atau yang lebih spesifiknya adala *text processing*. *Text Preprocessing* merupakan tahap yang sangat penting dalam *text mining*. Dalam *text mining*, *text preprocessing* merupakan suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur[10]. *Text preprocessing* memiliki beberapa tahapan, antara lain adalah *cleansing*, *case folding*, *tokenization*, normalisasi, *stopword removal*, dan *stemming*. *Cleansing* merupakan tahap untuk membersihkan data teks yang tidak diperlukan, antara lain menghapus nilai *null*, menghapus, nilai yang duplikat, menghapus karakter atau simbol yang tidak diperlukan dalam teks, dan memilih variable apa yang

akan digunakan. *Case Folding* adalah proses untuk memanipulasi teks merubah semua huruf kapital menjadi huruf kecil. *Tokenization* merupakan tahap memecah kalimat atau paragraf menjadi potongan kata tunggal atau disebut token. Normalisasi merupakan tahap merubah kata berbentuk singkatan, bahasa gaul, atau kata yang tidak baku menjadi kata yang baku. *Stopword removal* adalah tahap menghapus kata-kata yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna (*Stopword*)[11]. *Stemming* adalah proses mengganti kata yang memiliki imbuhan menjadi kata dasar[12].

Tahap selanjutnya setelah *data preprocessing* adalah pelabelan data. Data *tweet* yang dikumpulkan belum memiliki label yang menunjukkan *tweet* tersebut merupakan *tweet* yang bersifat positif atau negatif. Proses pelabelan data pada penilitian ini menggunakan salah satu library Python yaitu *VaderSentiment*. *VaderSentiment* merupakan salah satu library dalam Python yang akan memberi nilai apakah suatu kalimat itu bersifat positif atau negatif.

Tahap berikutnya adalah pembobotan kata menggunakan TF-IDF. TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan suatu metode untuk mengubah data dari data teks menjadi data numerik untuk dilakukan pembobotan pada tiap kata atau fitur. TF- IDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen. TF adalah frekuensi kemunculan kata dalam setiap dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam tiap dokumen tersebut. DF adalah frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. IDF adalah kebalikan dari nilai DF. Hasil dari pembobtan kata menggunakan TF-IDF ini adalah hasil perkalian dari TF dikalikan dengan IDF[13]. Pada TF-IDF rumus yang digunakan untuk menghitung bobot (W) dari masing-masing dokumen terhadap kata kunci adalah sebagai berikut:

$$W_{dt} = TF_{dt} IDF_{ft} \quad (1)$$

Dimana W_{dt} adalah nilai dokumen ke-d pada kata ke-t ; TF_{dt} adalah jumlah kata yang dicari dalam suatu dokumen ; IDF_{ft} adalah *Inverse Document Frequency* ($\log(\frac{N}{df})$) ; N adalah jumlah dokumen ; df adalah jumlah dokumen yang mengandung kata yang dicari.

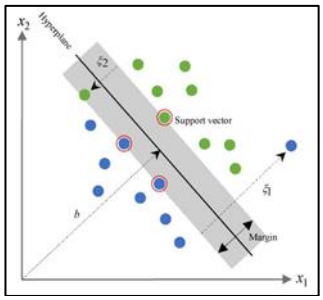
Setelah melakukan pembobotan, tahap selanjutnya adalah pemodelan klasifikasi menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. *Naïve Bayes* merupakan salah satu metode untuk melakukan klasifikasi. *Naïve Bayes* merupakan salah satu metode untuk melakukan klasifikasi. *Naïve Bayes* adalah teknik prediksi probabilitas sederhana berdasarkan pada penerapan Teorema Bayes [14]. Model yang digunakan dalam Naïve Bayes, adalah "model fitur independen"[15]. Algoritma *Naïve Bayes* didasarkan pada teorema bayes dengan rumus sebagai berikut :

$$P(H \mid E) = \frac{P(E|H) * P(H)}{P(E)} \quad (2)$$

Dimana P(H|E) adalah probabilitas akhir bersyarat suatu hipotesis H terjadi jika diberikan bukti E ; P(E|H) adalah probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H ; P(H) adalah probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun ; P(E) adalah probabilitas awal bukti E terjadi tanpa memandang hipotesis yang lain.

Support Vector Machine (SVM) adalah sebuah algoritma *supervised* untuk melakukan klasifikasi dengan membagi data menjadi dua kelas menggunakan garis vektor yang disebut *hyperplane*. Gambar 2 merupakan ilustrasi dari SVM.

Hyperplane terbaik diantara kedua kelas ditemukan dengan mengukur margin hyperplane dan mencari titik maksimal. Margin merupakan jarak antara *hyperplane* dengan pola terdekat dari setiap kelas. Pola yang paling dekat disebut sebagai support vector [16]. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM.



Gambar 2. Ilustrasi SVM

Tahap terakhir setelah melakukan pemodelan adalah melakukan evaluasi terhadap model dengan melihat akurasi dari model tersebut. Untuk melihat akurasi pada penelitian ini menggunakan *confusion matrix*. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya[17]. Gambar 3 merupakan tabel confusion matrix :

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Gambar 3. Tabel confusion matrix

Dimana TP adalah kelas yang diprediksi positif dan faktanya positif ; TN adalah kelas yang diprediksi negatif, dan faktanya negatif ; FP adalah kelas yang diprediksi positif dan faktanya negative ; FN adalah kelas yang diprediksi negatif dan faktanya positif. Nilai akurasi confusion matrix berdasarkan gambar 3 diperoleh dengan persamaan sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} * 100\% \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Pada penelitian ini data yang digunakan adalah tweet mengenai aplikasi PeduliLindungi yang diambil dari Twitter menggunakan library *Twint* di python. Data tweet tersebut nantinya akan dilakukan analisis sentimen dengan metode *Naïve Bayes* dan *Support Vector Machine* kemudian dilihat perbandingan akurasi dari dua metode tersebut. Data tweet yang diambil adalah tweet dari tanggal 2021-09-28 hingga 2021-10-07, jumlah data tersebut adalah 9476 dan memiliki 36 attribute antara lain id, *date*, *time*, *language*, *username*, *tweet*, dan sebagainya. Banyaknya data *tweet* yang digunakan adalah 5079 karena hanya *tweet* Bahasa Indonesia dan hanya attributes *tweet* yang dipakai pada penelitian ini. Data kemudian masuk ke tahap preprocessing untuk memperbaiki struktur dari data. Data *tweet* dapat dilihat pada tabel 1

Tabel 1. Data *Tweet*

	Tweet
1	Pasar Tradisional Kota Bandung Bakal Terapkan Aplikasi PeduliLindungi https://t.co/Cx5HGW3lGD
2	aplikasi peduli lindungi itu wajah dari developer gagal
3	Situs Palsu PeduliLindungi Minta Bayaran Rp 1 Juta untuk Pendaftaran Vaksin https://t.co/y12rcp9yQ2
4	Kementerian @kemkominfo mengimbau agar masyarakat hanya mengakses situs resmi PeduliLindungi serta mengunduh aplikasi resmi Peduli Lindungi yang tersedia di App Store dan Google Play Store.
5	Menkes Apresiasi Dukungan GoTo Sediakan Akses ke PeduliLindungi di Gojek dan Tokopedia https://t.co/uD2oXBNrKI https://t.co/ontTKAKChx
6	Sumpah ya benci bgt ditinggal kereta di depan mata grgr check in pedulilindungi nya lemot bgt tumben ga biasanyaðŸ˜- kan klo gini harus nunggu 30menit lg keretanya..... inginku berkata kasar _>\%]*Â£~'"!?<]#^
7	Pasar Tradisional Balubur Uji Coba Penerapan Aplikasi PeduliLindungi https://t.co/pOX3yxLzwL https://t.co/EjT6raONIP

3.2. Data *Pre-processing*

Data yang diperoleh dalam penelitian ini perlu dilakukan proses *pre-processing* agar menjadi lebih terstruktur. Terdapat beberapa tahapan dalam *data preprocessing*, tahap pertama adalah melakukan *cleansing data*. Pada tahap *cleansing*, nilai *null* dan nilai duplikat akan dihapus serta simbol atau karakter seperti alamat website, emoji, *mention* juga akan dihapus. Data yang melewati tahap *cleansing* kemudian akan dilakukan *case folding* dengan tujuan untuk mengubah semua huruf menjadi *lowercase* dan menghapus tanda baca agar data menjadi lebih terstruktur. Tahap berikutnya adalah melakukan *tokenization* pada data untuk memecah kalimat menjadi kata tunggal agar nantinya bisa dengan mudah melakukan normalisasi dan *stopword removal*. Setelah melakukan pemecahan kalimat menjadi kata tunggal tahap berikutnya adalah melakukan normalisasi dengan merubah kata singkatan atau kata tidak baku menjadi kata baku, contohnya seperti merubah kata "dgn" menjadi "dengan". Data yang sudah dinormalisasi kemudian akan dilakukan *stopword removal* yaitu menghapus kata-kata yang tidak berpengaruh terhadap proses kategorisasi. Contoh *stopword* dalam bahasa Indonesia adalah "yang", "dan", "di", "dari", dll. Proses ini bertujuan agar analisis dapat berfokus pada kata-kata yang penting saja. Setelah melewati tahap preprocessing jumlah data *tweet* menjadi 4636. Hasil dari proses *data pre-processing* dapat dilihat pada tabel 2

Tabel 2. Data *Tweet* setelah *Pre-processing*

	Tweet
1	pasar tradisional kota bandung terap aplikasi pedulilindungi
2	aplikasi peduli lindung wajah developer gagal
3	situs palsu pedulilindungi bayar rp juta daftar vaksin
4	menteri imbau masyarakat akses situs resmi pedulilindungi unduh aplikasi resmi peduli lindung sedia app store google play store
5	menkes apresiasi dukung goto sedia akses pedulilindungi gojek tokopedia
6	sumpah benci banget tinggal kereta mata grgr check in pedulilindungi nya lambat banget tumben ga kan gin nunggu menit kereta ingin kasar lt
7	pasar tradisional balubur uji coba terap aplikasi pedulilindungi

3.3. Pelabelan Data

Tahap pelabelan data dilakukan menggunakan *library VaderSentiment* di python. Sebelum menggunakan *VaderSentiment*, data *tweet* akan diterjemahkan terlebih dahulu menjadi bahasa Inggris karena *library VaderSentiment* hanya bisa menggunakan bahasa Inggris dalam menentukan apakah sifat dari *tweet* tersebut positif atau negatif. Hasil dari *VaderSentiment* berupa nilai antara -1 hingga 1, dimana nilai kurang dari 0 menunjukkan bahwa sentiment tersebut adalah negatif dan nilai lebih dari 0 menunjukkan bahwa sentimen tersebut adalah positif[18]. Berdasarkan nilai tersebut maka

dapat dilakukan pelabelan terhadap *tweet*. Hasil dari proses pelabelan data dapat dilihat pada tabel 3.

Tabel 3. Data *Tweet* dengan Label

	Tweet	Sentimen
1	pasar tradisional kota bandung terap aplikasi pedulilindungi	Positif
2	aplikasi peduli lindung wajah developer gagal	Negatif
3	situs palsu pedulilindungi bayar rp juta daftar vaksin	Negatif
4	menteri imbau masyarakat akses situs resmi pedulilindungi unduh aplikasi resmi peduli lindung sedia app store google play store	Positif
5	menkes apresiasi dukung goto sedia akses pedulilindungi gojek tokopedia	Positif
6	sumpah benci banget tinggal kereta mata grgr check in pedulilindungi nya lambat banget tumben ga kan gin nunggu menit kereta ingin kasar lt	Negatif
7	pasar tradisional balubur uji coba terap aplikasi pedulilindungi	Positif

Setelah melakukan pelabelan pada data, bisa dilihat dari jumlah data 4636 terdapat 4163 data yang memiliki sentimen positf dan 473 data yang memiliki sentimen negatif.

3.4. Pembobotan TF-IDF

Tahap ini merupakan tahap pembobotan kata menggunakan algoritma *Term Frequency – Invers Document Frequency* (TF-IDF), dimana setiap kata akan diberi nilai bobot sesuai dengan pengaruh atau seberapa sering kata tersebut muncul dalam suatu kalimat. Tahap pembobotan kata ini menggunakan *library* pada python yaitu *TFIdfVectorizer*. Nilai yang didapat dari pembobotan menunjukan jika semakin tinggi bobot kata terhadap suatu kalimat maka mengindikasikan bahwa kata tersebut semakin layak digunakan sebagai *keyword* terhadap kalimat tersebut. Contohnya bisa dilihat di tabel 4, pada Tweet1 bobot yang dimiliki kata “bandung” lebih besar dibandingkan kata “aplikasi”, maka bisa disimpulkan bahwa kata “bandung” bisa dijadikan salah satu keyword dari Tweet1.

Tabel 4. Pembobotan Kata

	akses	aplikasi	app	apresiasi	balubur	bandung	…….	vaksin	wajah
Tweet1	0	0.284991164	0	0	0	0.46263342	…….	0	0
Tweet2	0	0.282424277	0	0	0	0	…….	0	0.458466527
Tweet3	0	0	0	0	0	0	…….	0.380289056	0
Tweet4	0.190183014	0.141137718	0.229112455	0	0	0	…….	0	0
Tweet5	0.301029999	0	0	0.362649223	0	0	…….	0	0
Tweet6	0	0	0	0	0	0	…….	0	0
Tweet7	0	0.258652537	0	0	0.419877256	0	…….	0	0

3.5. Klasifikasi *Naïve Bayes*

Tahap yang dilakukan sebelum melakukan klasifikasi adalah membagai data menjadi data training dan data testing. Proses pembagian data menggunakan modul *train_test_split* dari *library sklearn.model_selection* di python. Rasio pembagian *data training* dan *data testing* adalah 70:30, dimana 70% untuk *data training* dan 30% untuk *data testing*. *Data training* digunakan untuk pembentuk model atau pola, sedangkan *data testing* digunakan untuk pengujian model. Data yang telah dibagi kemudian bisa digunakan untuk proses klasifikasi. Klasifikasi pertama dilakukan dengan algoritma *Naïve Bayes* menggunakan modul *Naïve Bayes* dari *library sklearn* di python. Hasil yang didapatkan dari klasifikasi *Naïve Bayes* bisa dilihat dari gambar 4. Hasil dari metode *Naïve Bayes* menyatakan bahwa tingkat akurasi pada metode ini adalah sebesar 90%, dimana nilai *precision negative* 10%, nilai *precision positive* 99%, nilai *recall negative* 68% dan nilai *recall positive* 91%.

```
Confusion matrix:
[[  15    7]
 [ 130 1239]]

Classification report:
              precision    recall  f1-score   support

    Negative       0.10      0.68      0.18        22
    Positive       0.99      0.91      0.95      1369

    accuracy                           0.90      1391
   macro avg       0.55      0.79      0.56      1391
weighted avg       0.98      0.90      0.94      1391
```

Gambar 4. Hasil Klasifikasi *Naïve Bayes*

3.6. Klasifikasi *Support Vector Machine*

Data yang telah dibagi kemudian bisa digunakan untuk proses klasifikasi. Klasifikasi pertama dilakukan dengan algoritma *Naïve Bayes* menggunakan modul *Naïve Bayes* dari *library sklearn* di python. Hasil yang didapatkan dari klasifikasi *Naïve Bayes* bisa dilihat dari gambar 5. Hasil dari metode *Support Vector Machine*

menyatakan bahwa tingkat akurasi pada metode ini adalah sebesar 91%, dimana nilai *precision negative* 16%, nilai *precision positive* 99%, nilai *recall negative* 74% dan nilai *recall positive* 91%.

```
Confusion matrix:
[[  23    8]
 [ 122 1238]]

Classification report:
              precision    recall  f1-score   support

    Negative       0.16      0.74      0.26        31
    Positive       0.99      0.91      0.95      1360

    accuracy                           0.91      1391
   macro avg       0.58      0.83      0.61      1391
weighted avg       0.97      0.91      0.93      1391
```

Gambar 5. Hasil Klasifikasi SVM

Hasil yang didapatkan dari klasifikasi Naïve Bayes dan SVM menunjukan bahwa SVM memiliki akurasi yang lebih tinggi dari pada Naïve Bayes dengan nilai akurasi SVM sebesar 91% dan nilai akurasi Naïve Bayes sebesar 90%. Penelitian yang dilakukan oleh Harun Sujadi, dkk dengan judul "Analisis Sentimen Pengguna Media Sosial *Twitter* terhadap Wabah Covid-19 dengan Metode *Naive Bayes Classifier* dan *Support Vector Machine*" , pada penelitian ini dilakukan analisis sentimen *Twitter* terhadap wabah *Covid-19* dengan menggunakan metode yang sama dengan penulis, yaitu *Naïve Bayes* dan SVM. Hasil akurasi yang didapatkan dari penelitian Harun Sujadi, dkk adalah 78.3% untuk metode *Naïve Bayes* dan 81.6% untuk metode SVM, dapat disimpulkan bahwa SVM lebih baik dibandingkan *Naïve Bayes* karena SVM memiliki nilai akurasi yang lebih tinggi [19]. Hasil yang diperoleh pada penelitian yang dilakukan oleh Harun Sujadi, dkk sama dengan hasil yang diperoleh pada penelitian penulis, dimana metode SVM lebih baik dibandingkan metode *Naïve Bayes*.

4. KESIMPULAN

Tujuan dilakukan penelitian ini untuk mengetahui hasil perbandingan tingkat akurasi dari metode klasifikasi yang digunakan yaitu *Naïve Bayes*, dan *Support Vector Machine* dengan menggunakan bahasa pemrograman python. Hasil yang didapatkan dari penelitian ini menunjukan nilai akurasi dari metode *Naïve Bayes* adalah adalah sebesar 90%, dengan nilai *precision negative* 10%, nilai *precision positive* 99%, nilai *recall negative* 68% dan nilai *recall positive* 91%. Nilai akurasi yang dari metode *Support Vector Machine* adalah 91%, dimana nilai *precision negative* 16%, nilai *precision positive* 99%, nilai *recall negative* 74% dan nilai *recall positive* 91%. Hasil yang didapat dari dua metode tersebut dalam penelitian dapat disimpulkan bahwa *Support Vector Machine* menghasilkan nilai akurasi yang lebih tinggi dari pada *Naïve Bayes* dengan perbedaan 1%. Dari hasil kesimpulan yang didapat, maka penelitian mengenai analisis sentimen dapat dikembangkan lagi seperti mencoba metode klasifikasi lainnya atau menggunakan *lexicon based*.

DAFTAR PUSTAKA

[1] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.

[2] N. Nurhidayati, S. Sugiyah, and K. Yuliantari, "Pengaturan Perlindungan Data Pribadi Dalam Penggunaan Aplikasi Pedulilindungi," *Widya Cipta: Jurnal Sekretari dan Manajemen*, vol. 5, no. 1, pp. 39–45, 2021, doi: 10.31294/widyacipta.v5i1.9447.

[3] S. Afrizal, H. N. Irmanda, N. Falih, and I. N. Isnainiyah, "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit," *Jurnal Informatik*, vol. 15, no. 3, pp. 157–168, 2019.

[4] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2580.

[5] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 2, pp. 1–9, 2021.

[6] B. Laurensz and Eko Sediyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.

[7] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Prisma*, vol. 4, pp. 628–634, 2021.

[8] Yudi Permana A and Makmun Effendi M, "Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN," *Jurnal Ilmiah Komputasi*, vol. 19, no. 1, pp. 39–50, 2020, doi: 10.32409/jikstik.19.1.154.

[9] W. Athira, I. Gholissodin, and rizal setya

Perdana, "Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.

[10] L. Hermawan and M. Bellaniar Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *Jurnal Transformatika*, vol. 17, no. 2, pp. 188–199, 2020, doi: 10.26623/transformatika.v17i2.1705.

[11] P. E. Mas'udia, M. D. Atmadja, and L. D. Mustafa, "INFORMATION RETRIEVAL TUGAS AKHIR DAN PERHITUNGAN KEMIRIPAN DOKUMEN MENGACU PADA ABSTRAK MENGGUNAKAN VECTOR SPACE MODEL," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 8, no. 1, pp. 355–362, 2017, doi: 10.24176/simet.v8i1.1016.

[12] A. Guterres, Gunawan, and J. Santoso, "Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based," *Teknika*, vol. 8, no. 2, pp. 142–147, 2019, doi: 10.34148/teknika.v8i2.224.

[13] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *Journal of Intelligent System and Computation*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.

[14] P. R. Sihombing and A. M. Arsani, "COMPARISON OF MACHINE LEARNING METHODS IN CLASSIFYING POVERTY IN INDONESIA IN 2018," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 1, pp. 51–56, 2021, doi: 10.20884/1.jutif.2021.2.1.52.

[15] B. Harijanto, Y. Ariyanto, and L. Miftahurroifa, "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI RETENSI ARSIP," *Jurnal Informatika Polinema*, vol. 4, no. 2, pp. 155–160, 2018, doi: 10.33795/jip.v4i2.159.

[16] A. Perdana and M. T. Furqon, "Penerapan Algoritma Support Vector Machine (SVM) Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia (Studi Kasus : RSJ . Radjiman Wediodiningrat , Lawang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 2, no. 9, pp. 3162–3167, 2018.

[17] S. Proboningrum and Acihmah Sidauruk, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SUPPLIER KAIN DENGAN METODE MOORA," *JSiI (Jurnal Sistem Informasi)*, vol. 8, no. 1, pp. 43–48, 2021, doi: 10.30656/jsii.v8i1.3073.

[18] P. Amira Sumitro *et al.*, "Analisis Sentimen Terhadapat Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based," *Jurnal Informatikan dan Teknologi Komputer*, vol. 2, no. 2, pp. 50–56, 2021, [Online]. Available: https://developer.twitter.com

[19] H. Sujadi, S. Fajar, and C. Roni, "ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE," *INFOTECH Journal*, vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.