

Terbit online pada laman web jurnal: <https://jurnal.plb.ac.id/index.php/tematik/index>

T E M A T I K

Jurnal Teknologi Informasi Komunikasi (e-Journal)

Vol. 11 No. 1 (2024) 56 - 64

ISSN Media Elektronik: 2443-3640

## Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine dan Lexicon Based

### *Sentiment Analysis of Used Clothing Import Ban Using Support Vectore Machine and Lexicon Methods*

Theresia Hendrawati<sup>1</sup>, Ni Luh Wiwik Sri Rahayu Ginantra<sup>2\*</sup>, Clarita Mutiara Saiman<sup>3</sup><sup>1,2,3</sup>Teknik Informatika, Fakultas Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia (INSTIKI)<sup>1</sup>theresia.hendrawati@instiki.ac.id, <sup>2</sup>wiwik@instiki.ac.id, claritamutiarasaiman@gmail.com.

#### Abstract

X as a social media platform, provides real-time tweets after the event, making it very relevant for sentiment analysis related to certain topics. This research discusses the prohibition on imports of used clothing in Indonesia using the SVM and Lexicon-based methods. The research aims to determine public sentiment regarding this government policy. The SVM method achieved 85.87% accuracy, with 93.83% recall for positive sentiment and 62.00% recall for negative sentiment. There were 76 wrong positive predictions with a precision of 88.11% and 37 wrong negative predictions with a precision of 77.02%. Meanwhile, the Lexicon Method achieved 60% accuracy, with a positive precision of 69% and a negative precision of 31%. Recall for the negative class is 25%, while for the positive class it is 75%. The results of sentiment analysis applying the Support Vector Machine method to build a classification model resulted in 753 data being successfully classified as positive sentiment, while 147 data were classified as negative sentiment. With different accuracies, it shows that sentiment analysis using the Support Vector Machine method has a higher level of accuracy than Lexicon.

Keywords: sentiment analysis, Twitter, SVM, lexicon-based

#### Abstrak

X sebagai platform media sosial, menyediakan tweet real-time setelah terjadinya peristiwa, menjadikannya sangat relevan untuk analisis sentimen terkait topik tertentu. Penelitian ini membahas larangan impor pakaian bekas di Indonesia dengan menggunakan metode SVM dan Lexicon Based. Tujuan penelitian adalah untuk mengetahui sentimen masyarakat terkait kebijakan pemerintah ini. Metode SVM mencapai akurasi 85.87%, dengan 93.83% recall untuk sentimen positif dan 62.00% recall untuk sentimen negatif. Terdapat 76 prediksi positif yang salah dengan precision 88.11%, serta 37 prediksi negatif yang salah dengan precision 77.02%. Sedangkan Metode *Lexicon Based* mencapai akurasi 60%, dengan precision positif 69% dan precision negatif 31%. Recall untuk kelas negatif adalah 25%, sedangkan untuk kelas positif adalah 75%. Hasil analisis sentimen penerapan metode Support Vectore Machine untuk membangun model klasifikasi menghasilkan 753 data berhasil diklasifikasikan sebagai sentimen positif, sementara 147 data diklasifikasikan sebagai sentimen negatif. Nilai akurasi yang berbeda menunjukkan bahwa, analisis sentimen menggunakan metode Support Vectore Machine memiliki tingkat akurasi yang lebih tinggi dari pada Lexicon Based.

Kata kunci: analisis sentimen, twitter, SVM, lexicon-based.

#### 1. Pendahuluan

Di era modern ini, teknologi sangat erat kaitannya dengan internet. Perkembangan teknologi dan informasi memberikan dampak positif bagi masyarakat Indonesia. Seperti halnya informasi masyarakat bisa mendapatkan dan memberikan informasi melalui media sosial. Media sosial adalah platform digital yang memungkinkan pengguna untuk berinteraksi, berbagi, bekerja sama, berkomunikasi

dengan pengguna lain membentuk ikatan sosial secara virtual melalui internet [1], [2]. Saat ini masyarakat banyak menggunakan media sosial untuk berbagi pendapat atau opini terhadap suatu isu yang sedang populer di kalangan masyarakat. Media sosial seperti Facebook, Tiktok, X, Instagram dan Youtube merupakan media yang sangat populer dikalangan masyarakat saat ini.

Telset.id menyatakan bahwa X merupakan situs media sosial yang memiliki pengguna aktif yang masuk ke dalam 10 aplikasi media sosial terpopuler di dunia [3]. Dimana Indonesia menduduki peringkat ke empat dengan jumlah 78 juta pengguna X aktif. Aplikasi X banyak memberikan informasi berupa data teks, opini pengguna serta pendapat tertulis dalam kalimat-kalimat yang berupa teks dan tidak terstruktur. X adalah platform media sosial yang menyajikan data real-time berupa tweet yang diunggah segera setelah peristiwa terjadi [4]. Kecepatan ini membuat X sangat relevan untuk analisis sentimen real-time terkait topik atau peristiwa tertentu. Data di X bersifat terbuka dan sebagian besar tweet bersifat publik, memungkinkan peneliti untuk mengumpulkan data besar secara langsung tanpa izin khusus [5]. Jumlah pengguna aktif yang besar, X menciptakan volume besar tweet setiap harinya, menghasilkan dataset yang luas dan beragam untuk analisis sentimen. Selain itu, X sering digunakan oleh individu, tokoh terkenal, perusahaan, dan media, sehingga peristiwa besar dapat memiliki dampak langsung pada opini publik, menjadikannya sumber informasi berharga untuk analisis sentimen [6]. Melihat respon masyarakat terhadap kebijakan pemerintah tentang larangan impor pakaian bekas dapat dilihat melalui sarana dan media sosial salah satunya melalui X. Alasan peneliti tertarik melakukan analisis sentimen mengenai larangan impor pakaian bekas karena pembahasan mengenai kebijakan pemerintah ini tengah menjadi trending topic di X.

Larangan impor pakaian bekas merupakan kebijakan yang kontroversial dimasyarakat saat ini. Apalagi saat ini trend thrift dan bisnis pakaian bekas bermerek masih berlanjut dari tahun ketahun. Pakaian thrift menjadi salah satu alternatif yang masih disukai oleh para pemuda dan kaum millennial. Diterbitkannya kebijakan Pemerintah tentang larangan impor baju bekas yang tertulis dalam Permendag Nomor 40 tahun 2022 tentang perubahan atas Permendag Nomor 18 tahun 2021 tentang Barang Dilarang Ekspor dan Barang Dilarang Impor [7], maka impor baju bekas dilarang di Indonesia. Kebijakan ini dibuat karena baju bekas memiliki potensi membawa resiko kesehatan bagi masyarakat. Selain itu, baju bekas juga dapat mengancam industri tekstil dan produk fashion dalam negeri. Larangan impor baju bekas, pemerintah berharap dapat mendorong pertumbuhan industri tekstil dan produk fashion dalam negeri serta meningkatkan perlindungan konsumen terhadap produk yang tidak aman dan berisiko.

Kebijakan tersebut memiliki dampak signifikan pada masyarakat, terutama pada konsumen dan pengusaha thrift. Konsumen akan merasakan dampak negatif karena penurunan persaingan pasar menyebabkan harga pakaian baru cenderung lebih mahal. Pengusaha thrift juga menghadapi kesulitan dalam menjaga bisnis mereka karena sulitnya mencari persediaan pakaian

bekas dan meningkatnya harga pakaian bekas. Selain itu, masyarakat yang bergantung pada impor pakaian bekas untuk memenuhi kebutuhan mereka juga dapat terpengaruh karena ketersediaan pakaian yang terjangkau menjadi lebih sedikit.

Berkaitan dengan judul penelitian yang diteliti mengenai analisis sentimen larangan impor pakaian bekas menggunakan metode SVM dan Lexicon Based, terdapat beberapa referensi terkait, diantaranya pada penelitian yang dilakukan oleh M. A. Sobirin [8] tentang pemberlakuan Karantina dimasa covid-19 di Indonesia dengan metode *Support Vector Machine* (SVM) dan Lexicon Based. Hasil penelitian ini menunjukkan Algoritma *Support Vector Machine* (SVM) dengan pelabelan manual menghasilkan hasil tertinggi dengan percobaan oversampling menggunakan skenario ke-5 yaitu 90% data training dan 10% data testing dengan hasil akurasi sebesar 97,36%, presisi 97%, recall 98%, dan f1-score 97% sedangkan hasil pelabelan menggunakan InSetLexicon menghasilkan hasil tertinggi dengan percobaan oversampling menggunakan skenario ke-5 juga yaitu 90% data training dan 10% data testing dengan hasil akurasi sebesar 97%, presisi 96%, recall 96%, dan f1-score 97%. Pada penelitian Mahendrajaya dkk [9] tentang pengguna Gopay Menggunakan Metode *Lexicon Based* dan *Support Vector Machine*. Dari penelitian ini juga menunjukkan metode klasifikasi SVM dengan membandingkan kernel cukup baik. Kernel linear mendapatkan akurasi yang cukup baik sebesar 89,17% dimana sistem kernel dapat melakukan klasifikasi dengan sebanyak 1109 ulasan. Sedangkan untuk kernel polynomial mendapatkan akurasi sebesar 84,38% dan dapat melakukan klasifikasi dengan benar, sebanyak 1021 ulasan. Penelitian ini juga dapat disimpulkan bahwa data twitter untuk ulasan *Go-Pay* dapat dipisah secara linear cukup baik. Penelitian lainnya tentang *Support Vector Machine* terkait analisis sentimen pengguna sistem *PayLater* menggunakan metode pembobotan *Lexicon* dilakukan oleh Wahidna dan Nerisafitra [10]. Penelitian ini menunjukkan hasil pengujian penerapan metode SVM dengan 50,4% netral, 38,8% negatif, dan 10,8% positif pada *Shopee Paylater*. Sedangkan, pada data *Go Paylater*, 47,8% diklasifikasikan sebagai netral, 33,5% negatif, dan 18,7% positif. Hal ini menunjukkan bahwa sentimen pengguna *Shopee paylater* atau *Go paylater* cenderung netral berdasarkan klasifikasi oleh leksikon. Penelitian dengan metode SVM dan *Lexicon Based* sebelumnya dilakukan oleh Zurriel dan Fahrurrozi [11] tentang klasifikasi sentiment terhadap kebijakan Psbb. Hasil pengukuran performa model klasifikasi SVM dengan kernel RBF mendapatkan hasil keseluruhan yang paling baik dibanding model dengan kernel lainnya. Nilai akurasi, precision, recall, dan f1-score yang dimiliki oleh model tersebut secara berturut-turut adalah 95.94%, 94.41 %, 97.8%, dan 96.08%. Model

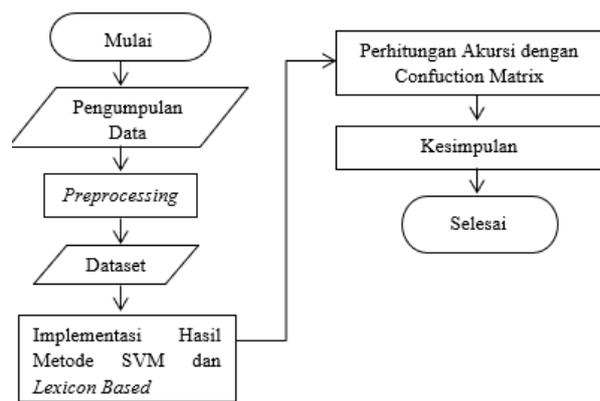
klasifikasi memprediksi terdapat sentimen positif sebanyak 11.764 tweet (52.7%) dan sentimen negatif sebanyak 10.571 tweet (47.3%). Dari hasil ini disimpulkan bahwa dari 22.335 tweet pengguna *Twitter* yang diambil pada tanggal 28 Mei, 16 September dan 27 September 2020, cenderung bersentimen positif terhadap kebijakan PSBB di Indonesia. Penelitian lainnya yang dilakukan oleh V.A.R.Barao dkk yang juga menggunakan metode *Naïve Bayes* dan SVM dengan *Lexicon Based* untuk analisis sentimen pada *twitter*. Penelitian ini menghasilkan menghasilkan nilai akurasi, *presisi*, dan *recall* sebesar 74,44%, 56,84%, 51,84% pada metode *Naïve Bayes*. Pada metode *Support Vector Machine* menguji tiga kernel yaitu: linear, rbf dan poly dimana kernel terbaik adalah kernel rbf dengan nilai parameter terbaik dari  $C=100$  dan  $\gamma=0,01$  dimana hasil akurasi, *presisi*, dan *recall* sebesar 85,28%, 75,96%, 76,36% [12].

Penelitian yang dilakukan sebelumnya, peneliti menerapkan beberapa aspek untuk digunakan dalam penelitian saat ini agar dapat membedakannya dari penelitian sebelumnya. Peneliti akan membandingkan dua metode, yaitu Metode SVM dan *Lexicon Based*, untuk menganalisis dampak larangan impor pakaian bekas di Indonesia. Alasan peneliti memilih kedua metode ini karena SVM telah terbukti efektif dalam mengklasifikasikan dokumen dalam klasifikasi teks, termasuk analisis sentimen. SVM dapat memisahkan data secara jelas berdasarkan fitur nilai, dan dapat menangani data dengan dimensi yang tinggi, yang sering terjadi pada data teks. SVM juga menunjukkan akurasi yang lebih tinggi dalam klasifikasi sentimen dibandingkan dengan algoritma lainnya. Sementara metode *Lexicon Based* dapat membantu mengidentifikasi sentimen positif atau negatif yang terdapat dalam teks. Dalam kasus ini, *lexicon* akan berisi kumpulan kata positif dan negatif yang digunakan untuk menandai sentimen dalam teks, data kalimat sentimen akan lebih mudah untuk diklasifikasikan [9].

Dengan mengetahui nilai akurasi dua metode yaitu metode SVM dan *Lexicon Based* dalam analisis sentimen terkait larangan impor pakaian bekas, diharapkan dapat memperoleh hasil yang lebih akurat dalam mengevaluasi pandangan dan opini masyarakat tentang kebijakan ini. Hasil analisis sentimen ini dapat menjadi acuan bagi pemerintah dalam memutuskan kebijakan yang lebih baik dan tepat terkait dengan larangan impor pakaian bekas.

## 2. Metode Penelitian

Dalam penelitian ini, dilakukan beberapa langkah atau proses untuk mendapatkan hasil dan kesimpulan yang didapat dari penelitian yang dilakukan. Gambar 1 menggambarkan alur penelitian yang dilakukan:



Gambar 1. Alur Penelitian (dokumentasi pribadi)

Pengumpulan data dilakukan dengan dua tahap yaitu crawling dan labeling. Proses crawling dilakukan dengan mengambil tweet dari *Twitter* menggunakan aplikasi rapid miner, dan data yang dikumpulkan terdiri dari 1000 tweet. Tweet yang diambil dipilih berdasarkan isu yang sedang viral yaitu larangan impor pakaian bekas. Selanjutnya, tahap labeling dilakukan untuk memberikan label atau kategori pada data, dengan label yang diberikan adalah positif atau negatif.

*Preprocessing teks* atau pra- pemrosesan teks adalah proses persiapan data teks sebelum dilakukan analisis atau pemrosesan lebih lanjut. *Preprocessing teks* meliputi beberapa tahapan seperti membersihkan teks dari karakter yang tidak diperlukan, mengubah format teks menjadi lebih muda diposes, dan melakukan normalisasi teks agar lebih konsisten [13], [14]. Menurut [15], *preprocessing teks* adalah proses perubahan bentuk sebuah data menjadi lebih terstruktur sesuai dengan kebutuhan dalam proses data mining disebut sebagai preprocessing data. Tujuan utamanya adalah menghilangkan noise dan membuat data teks menjadi lebih mudah dimengerti oleh algoritma pembelajaran mesin.

Pelabelan kelas opini atau sentimen dimana data yang telah dikumpulkan dilabeli ke dalam kelas positif atau negatif [16]. Dalam penelitian ini menggunakan algoritma *Lexicon Based* untuk menentukan daftar kata positif dan daftar kata negatif. Proses melakukan pelabelan tersebut, digunakan vader *lexicon*. Pada tahap ini data dilabelkan menjadi kelas positif dan negatif berdasarkan perhitungan atau skoring yang dilakukan. Setiap kata positif yang terkandung pada suatu tweet maka skor +1 akan ditambahkan pada tweet tersebut. Sebaliknya untuk setiap kata negatif yang terkandung pada suatu tweet maka skor -1 akan dikurangkan pada tweet tersebut. Skor akhir dihitung dengan menjumlahkan seluruh skor. Jika diperoleh total skor  $>0$  maka tweet diklasifikasikan ke dalam kelas positif. Sebaliknya, jika diperoleh total skor  $<0$  maka tweet diklasifikasikan ke dalam kelas negative.

SVM adalah metode pembelajaran mesin yang sangat berguna dalam memecahkan masalah klasifikasi, regresi, dan pengenalan pola [17]. SVM dapat digunakan untuk memisahkan data yang tidak linear dan memiliki banyak fitur, serta memiliki keunggulan dalam menghindari overfitting [18]. SVM pertama kali diperkenalkan oleh Boser, Guyon dan Vapnik pada tahun 1992 di Annual Workshop on Computational Learning Theory. Metode *Support Vector Machine* adalah salah satu teknik pembelajaran mesin (*machine learning*) yang digunakan untuk memisahkan data ke dalam kelas-kelas yang berbeda dengan menggunakan fungsi kernel [19][20]. Pada penelitian ini SVM digunakan untuk membangun model klasifikasi yang dapat membedakan antara teks yang mengandung sentimen positif atau negatif terkait dengan kebijakan larangan impor pakaian bekas. Pendekatan SVM digunakan untuk mendapatkan akurasi terbaik. Dalam metode pemodelan ini menggunakan operator SVM atau LibSVM yang disediakan oleh RapidMiner.

Metode *Lexicon Based* adalah teknik yang menggunakan daftar kata-kata dengan sentimen tertentu untuk mengklasifikasikan teks ke dalam kelas sentimen yang berbeda berdasarkan kata positif, negative ataupun netral [21][22]. Metode ini dapat digunakan untuk analisis sentimen pada berbagai jenis teks, seperti ulasan produk, media sosial, atau artikel berita [23]. Dalam analisis sentimen larangan impor pakaian bekas, metode *Lexicon Based* digunakan untuk menghitung skor sentimen pada setiap dokumen teks berdasarkan jumlah kata positif atau negatif yang terkandung dalam dokumen tersebut. Proses pembobotan dalam metode *Lexicon Based* dilakukan dengan mengalokasikan skor tertentu pada setiap kata dalam kamus berdasarkan kategori sentimennya [24].

Pada analisis sentimen larangan impor pakaian bekas menggunakan metode SVM dan lexicon-based, pengujian dilakukan untuk mengevaluasi kinerja model analisis sentimen yang telah dibangun. Pada tahap pengujian, beberapa metrik digunakan untuk mengukur kinerja model, seperti akurasi, presisi, recall, dan F1-score. Akurasi mengukur persentase prediksi yang benar secara keseluruhan, sedangkan presisi dan recall mengukur kinerja model dalam memprediksi sentimen positif dan negatif. F1-score merupakan rata-rata harmonis antara presisi dan recall. Pada penelitian ini untuk menguji tingkat akurasi dari algoritma SVM dan *Lexicon Based* peneliti menggunakan confusion matrix. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Adapun persamaan dalam melakukan evaluasi Confusion Matrix untuk menghitung Accuracy (Rumus 1), Precision (Rumus 2) dan Recall (Rumus 3) yaitu:

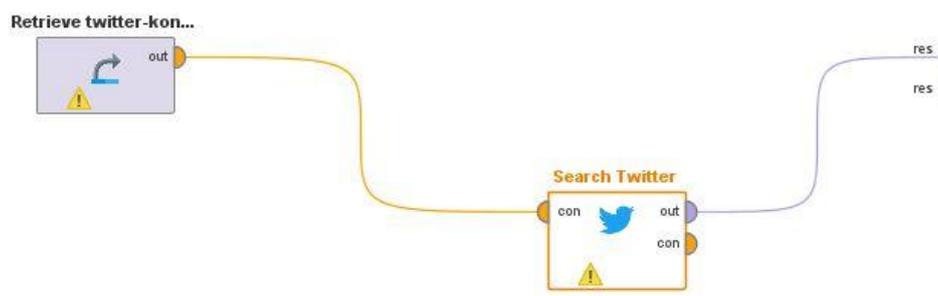
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

### 3. Hasil dan Pembahasan

Dalam penelitian ini, sebanyak 5000 data dari X berhasil dikumpulkan. Data yang dikumpulkan berupa data teks yang diambil dengan teknik crawling pada X. Data yang terkumpul akan mengalami proses pra-pemrosesan untuk memperoleh data yang terstruktur dan siap digunakan. Setelah pra-pemrosesan selesai, diperoleh 1000 data yang bersih dan siap digunakan dalam penelitian.



Gambar 2. Crawling Data

Tahap *Preprocessing* dalam pengolahan data melibatkan serangkaian proses untuk membersihkan, memformat, dan mempersiapkan data mentah agar sesuai untuk analisis dan pemrosesan lebih lanjut [25]. Dalam tahap ini, dilakukan pembersihan data dengan menghapus beberapa tweet yang tidak sesuai dengan topik penelitian secara manual dan menghilangkan tweet yang sama dengan menggunakan operator

remove duplicates. Tahapan preprocessing dimulai dengan memisahkan kata dengan kata lainnya menggunakan tokenize, mengubah seluruh kalimat menjadi lower case, kemudian memfilter seluruh kata dengan menggunakan filter stopwords dan filter tokens. Sehingga, dari 5000 data didapatkan 1000 dataset yang digunakan dalam penelitian ini. Tahapan preprocessing dapat digambarkan pada Gambar 3.



Gambar 3. Tahap *Preprocessing* (dokumentasi pribadi)

Tabel 1. Hasil Pelabelan (dokumentasi pribadi)

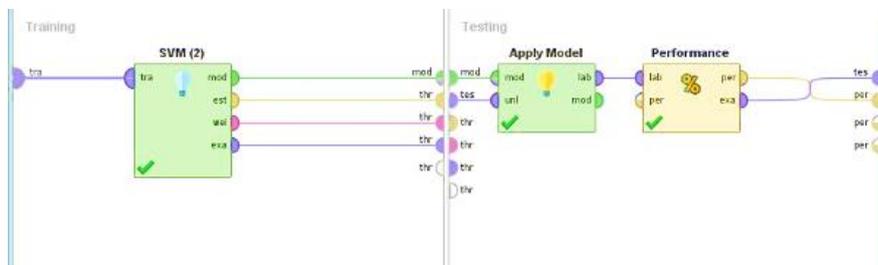
Tweet	Compound Score	Sentiments
the government has firmly also banned imports of used clothing	-0.4588	Negatives

Proses pelabelan dilakukan dengan menginstal *library Vader sentiment* menggunakan kode program `!pip install VaderSentiment` dan mengimpor *library SentimentIntensityAnalyzer* pada google collab.

Kemudian menghitung *sentiment scores* dari teks yang terdapat dalam kolom Tweet pada DataFrame `tweet_df` menggunakan sebuah objek analisis sentimen yang disebut *analyser*. Selanjutnya akan dilakukan proses penentuan kategori sentimen berdasarkan *compound scores*.

Data tweet yang  $< 0$  akan dilabelkan ke dalam kategori negatif, sedangkan data tweet yang  $> 0$  akan dilabelkan ke dalam kategori positif seperti terlihat pada Tabel 1.

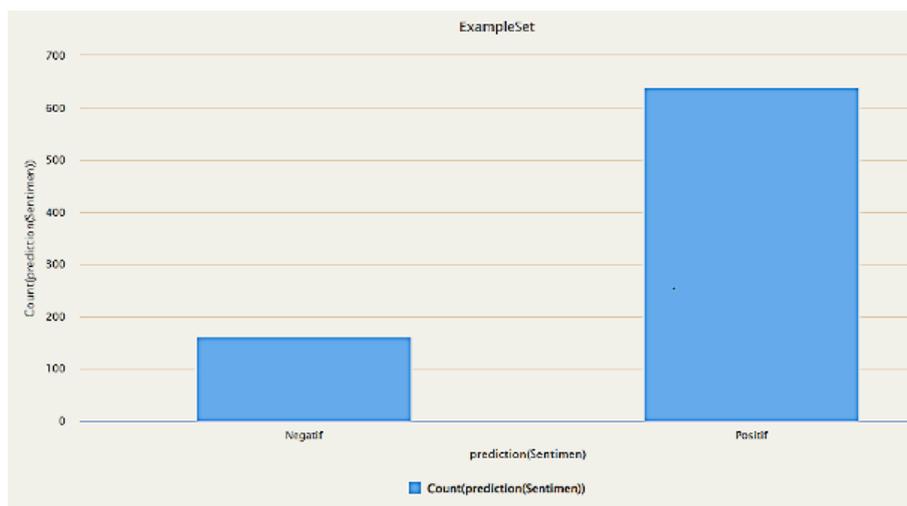
Skema yang digunakan pada metode Support Vector Machine (SVM) digambarkan pada Gambar 4.



Gambar 4. Skema SVM

Penerapan metode *Support Vector Machine* pada Gambar 4 dimulai dengan operator SVM yang berfungsi untuk menerapkan data ke dalam algoritma SVM, kemudian menerapkan model yang telah dilatih sebelumnya menggunakan *apply model* dan langkah

terakhir adalah menggunakan *performance* untuk mengevaluasi hasil kinerja model yang berupa *accuracy*, *precision* dan *recall* yang ditampilkan secara visual pada Gambar 5.



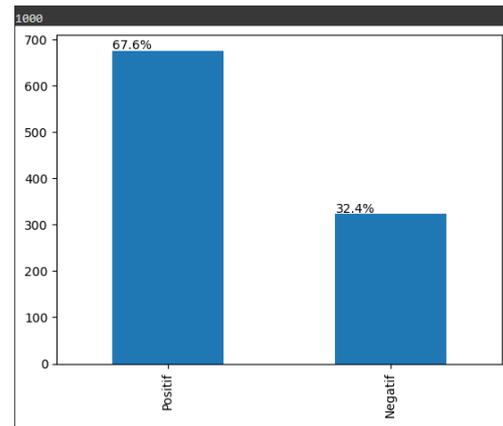
Gambar 5. Hasil Prediksi SVM (dokumentasi pribadi)

Dalam menerapkan metode *Lexicon Based* dengan menggunakan google colab dan bahasa pemrograman Python dalam pengimplementasiannya. Proses selanjutnya yaitu penggunaan

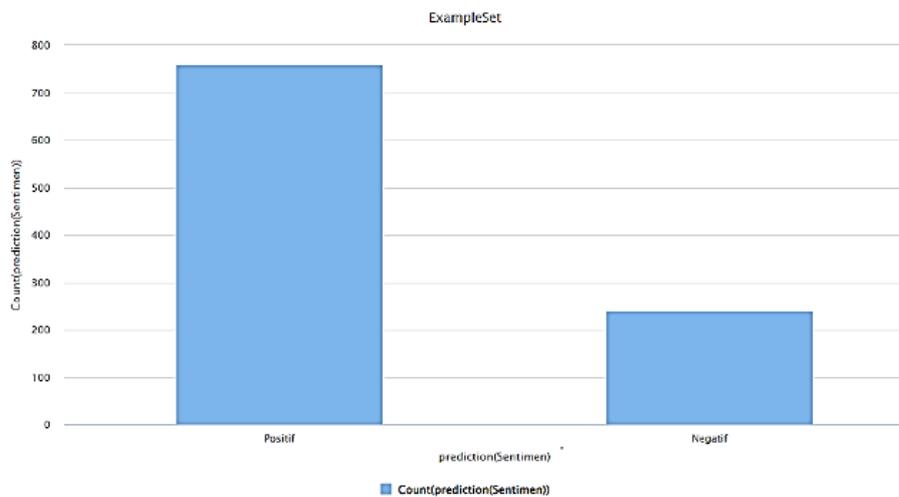
*SentimentIntensityAnalyzer* dari pustaka *vaderSentiment* untuk melakukan analisis sentimen pada teks.

Grafik hasil prediksi *Lexicon Based* menampilkan 690 data dengan presentase 67.6% bersentimen positif dan 310 data dengan presentase 32.4% bersentimen negatif yang ditampilkan secara visual pada Gambar 6.

Metode *Support Vector Machine* digunakan sebagai algoritma untuk membangun klasifikasi model. Dataset yang digunakan dalam proses ini adalah 1000 data dimana, 300 dataset pertama yang sudah dilabeli digunakan untuk membangun model. Selanjutnya peneliti akan melakukan pelabelan otomatis terhadap 1000 dataset menggunakan *rapidminer*, dengan memanfaatkan model yang telah dibangun sebelumnya.



Gambar 6. Hasil Prediksi *Lexicon Based* (dokumentasi pribadi)



Gambar 7. Hasil Klasifikasi Sentimen (dokumentasi pribadi)

Dari 1000 dataset yang digunakan, maka diperoleh hasil untuk klasifikasi sentimen berupa 753 data berhasil diklasifikasikan sebagai sentimen positif, dan 247 data diklasifikasikan sebagai sentimen negative. Grafik visualisasi hasil klasifikasi ditampilkan pada Gambar 7.

### 3.1. Pembahasan

Dari hasil analisis sentimen terhadap larangan impor pakaian bekas menggunakan metode SVM dan *Lexicon Based*, menunjukkan tingkat akurasi yang berbeda. Hasil evaluasi kinerja model yang diperoleh pada metode SVM dengan total sentimen sebanyak 1000 yang ditampilkan *Confusion matrix* pada tabel 2 sebagai berikut:

Tabel 2. *Confusion Matrix* Metode SVM (dokumentasi pribadi)

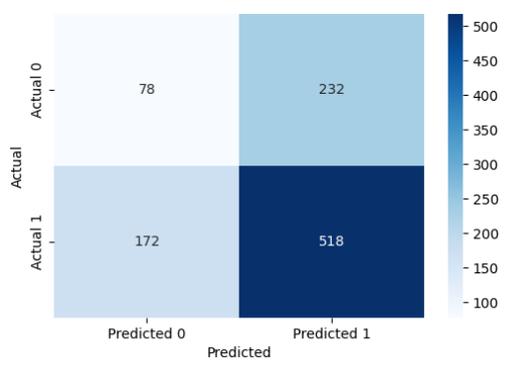
	True Negatif	True Positif	Class Precision
Pred.Negatif	124	37	77.02%
Pred.Positif	76	563	88.11%
Class Recall	62%	93.83%	

Jumlah data yang benar-benar mengandung sentimen positif dan berhasil diklasifikasikan dengan benar oleh model sebanyak 563 data dengan class recall positif

93.83%. Kasus negatif dan diprediksi benar negatif oleh model adalah 124 data dengan class recall 62.00%. Proporsi kasus prediksi positif yang memprediksi kasus yang sebenarnya negatif sebanyak 76 data dengan kelas precision positif 88.11%. Proporsi kasus prediksi negatif yang memprediksi kasus yang sebenarnya positif sebanyak 37 data dengan class precision negatif 77.02%. Hasil *Confusion matrix* maka memperoleh akurasi sebesar 85.87%. Sedangkan hasil evaluasi algoritma *Lexicon Based* ditampilkan pada *Confusion matrix* tampak pada Gambar 8.

Dari 1000 data tweet yang dianalisis menggunakan *Confusion matrix*, terdapat 518 prediksi benar positif dan 78 prediksi benar negatif. Namun, terdapat juga 172 kasus di mana kelas negatif diprediksi sebagai positif oleh model, dan sebaliknya, terdapat 232 kasus di mana kelas positif diprediksi sebagai negatif oleh model tampak pada gambar 8. Terdapat dua kelas yang dievaluasi, yaitu negatif dan positif. Precision, yang mengukur akurasi prediksi positif, memiliki nilai 0.31 (31%) untuk kelas negatif dan 0.69 (69%) untuk kelas positif. Recall, yang mengukur kemampuan model dalam mendeteksi kasus positif, adalah 0.25 (25%)

untuk kelas negatif dan 0.75 (75%) untuk kelas positif. Jumlah sampel atau data yang termasuk dalam setiap kategori juga tertera. Akurasi keseluruhan model adalah 60%. Hasil evaluasi model *Lexicon Based* dapat dilihat pada Tabel 3.



Gambar 8. Confusion matrix Metode *Lexicon Based* (dokumentasi pribadi)

Tabel 3. Evaluasi Model *Lexicon Based* (dokumentasi pribadi)

	Precision	Recall	F1-Score	Support
Negatif	0.31	0.25	0.28	310
Positif	0.69	0.75	0.72	690
Accuracy			0.60	1000

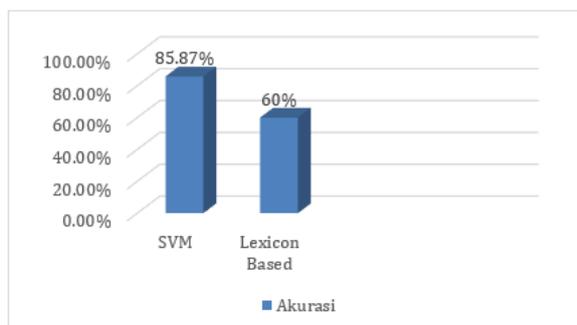
### 3.2. Perbandingan Hasil Analisis Sentimen

Dari hasil implementasi metode *Support Vector Machine* dan *Lexicon Based*, menggunakan 1000 data dan menghasilkan evaluasi model yang berbeda-beda. Berikut perbedaan setiap evaluasi model dari kedua metode ditampilkan pada Tabel 4.

Tabel 4. Perbandingan *Accuracy* (dokumentasi pribadi)

Metode	Accuracy
<i>Support Vector Machine</i>	0.31
<i>Lexicon Based</i>	0.69

Berdasarkan data pada Tabel 4, dapat disimpulkan bahwa metode *Support Vector Machine* menunjukkan tingkat akurasi yang lebih tinggi dibandingkan dengan metode *Lexicon Based*, yakni sekitar 85.87% untuk *Support Vector Machine* dan sekitar 60% untuk metode *Lexicon Based* dan hasil visualisasinya digambarkan pada Gambar 9.

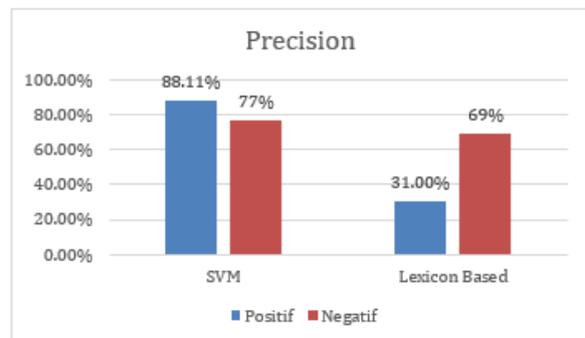


Gambar 9. Grafik Perbandingan *Accuracy* (dokumentasi pribadi)

Dari hasil yang tertera pada Tabel 5, dapat dilihat bahwa metode *Support Vector Machine* menunjukkan tingkat presisi yang lebih tinggi untuk kedua kelas, positif dan negatif, dibandingkan dengan metode *Lexicon Based*. Nilai presisi sekitar 88.11% untuk kelas positif dan 77.02% untuk kelas negatif, SVM menunjukkan kecenderungan untuk memberikan prediksi yang lebih akurat. Sementara itu, metode *Lexicon Based* memiliki presisi sekitar 31% untuk kelas positif dan 69% untuk kelas negatif dan digambarkan secara visual pada Gambar 10. Hal ini menunjukkan bahwa metode *Support Vector Machine* lebih handal dalam mengklasifikasikan sentimen dalam kasus ini

Tabel 5. Perbandingan *Precision* (dokumentasi pribadi)

Metode	<i>Precision</i>	
	Positif	Negatif
<i>Support Vector Machine</i>	88.11%	77.02%
<i>Lexicon Based</i>	31%	69%

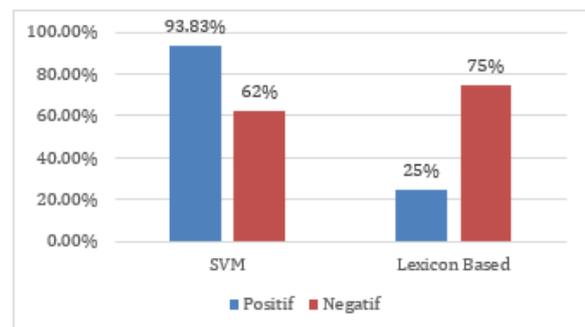


Gambar 10. Grafik Perbandingan *Precision* (dokumentasi pribadi)

Tabel 6. Perbandingan *Recall* (dokumentasi pribadi)

Metode	<i>Recall</i>	
	Positif	Negatif
<i>Support Vector Machine</i>	93.83%	62.00%
<i>Lexicon Based</i>	25%	75%

Dari hasil yang tertera pada Tabel 6, terlihat bahwa metode *Support Vector Machine* memiliki tingkat *recall* yang lebih tinggi untuk kelas positif dengan persentase sekitar 93.83%, menunjukkan kemampuannya dalam mendeteksi kasus positif dengan akurat tampak pada Gambar 11.



Gambar 11. Grafik Perbandingan *Recall* (dokumentasi pribadi)

Di sisi lain, untuk kelas negatif, metode *Lexicon Based* memiliki *recall* sekitar 75%, menandakan kemampuannya dalam mendeteksi kasus negatif secara efisien. Kedua metode ini menonjol dalam aspek tertentu. SVM unggul dalam mendeteksi kasus positif, sementara pendekatan berbasis *Lexicon Based* lebih handal dalam mendeteksi kasus negatif.

### 3.3. Perbedaan Hasil Klasifikasi Sentimen

Penelitian ini mengambil dua contoh sampel perbedaan prediksi sentimen pada kedua metode yang digunakan. Tabel 7 menunjukkan perbedaan yang muncul antara kedua metode, yakni *Support Vector Machine* dan *Lexicon Based*, dalam menentukan sentimen positif dan negatif dari suatu tweet.

Tabel 7. Perbedaan Prediksi Sentimen (dokumentasi pribadi)

Teks	Prediksi SVM	Prediksi Lexicon Based
pemerintah turut tegas dengan melarang impor pakaian bekas	Positif	Negatif
jangan sampai industri tekstil dalam negerikalalah dengan impor pakaian bekas	Positif	Negatif

Perbedaan antara metode *Support Vector Machine* (SVM) dan *Lexicon Based* dalam menentukan sentimen positif dan negatif dari tweet dapat diamati dalam Tabel 7. Dalam contoh pertama, SVM memprediksi sentimen positif terkait dengan keputusan tegas pemerintah melarang impor pakaian bekas, sedangkan *Lexicon Based* memprediksi sentimen negatif. Pola kalimat pertama ini mencerminkan tindakan tegas pemerintah dalam melarang impor pakaian bekas. SVM *menginterpretasikan* kata-kata seperti "turut tegas" dan "melarang" sebagai tindakan positif dan kuat dari pemerintah. SVM mampu menangkap pola dan nuansa yang menyiratkan keputusan tegas sebagai sentimen positif, dan dengan demikian, kalimat ini kemungkin diklasifikasikan sebagai sentimen positif. Sedangkan *Lexicon Based* mungkin menghadapi keterbatasan dalam daftar kata yang telah ditentukan sebelumnya. Kata-kata seperti "tegas" atau "melarang" mungkin tidak selalu terdaftar sebagai sentimen positif dalam *lexicon*. Hal ini dapat menyebabkan metode ini menginterpretasikan kalimat sebagai sentimen negatif.

Dalam contoh kedua, SVM memprediksi sentimen positif terkait dengan ketidaksetujuan terhadap impor pakaian bekas agar industri tekstil dalam negeri tidak kalah, sementara *Lexicon Based* memprediksi sentimen negatif. SVM dapat mengenali nuansa positif dalam kalimat tersebut, menginterpretasikan keinginan untuk menjaga keberlanjutan industri tekstil sebagai sentimen positif. Oleh karena itu, kalimat ini

diklasifikasikan sebagai sentimen positif oleh SVM. Sementara *Lexicon Based* mungkin menghadapi keterbatasan dalam daftar kata yang telah ditentukan sebelumnya. Kata-kata seperti "jangan sampai kalah" mungkin terdaftar sebagai kata negatif dalam *lexicon* sehingga menghasilkan interpretasi negatif terhadap kalimat tersebut.

Perbedaan dalam interpretasi sentimen antara kedua metode mencerminkan sensitivitas metode terhadap kompleksitas makna kalimat dan keterbatasan *Lexicon Based* dalam menghadapi variasi dalam teks yang tidak selalu tercakup dalam daftar kata atau aturan tertentu. SVM lebih unggul daripada *Lexicon Based* dalam analisis sentimen terkait kebijakan larangan impor pakaian bekas. Kelebihan SVM sebagai model pembelajaran mendalam memungkinkan penangkapan pola yang kompleks, sementara adaptasinya terhadap variasi bahasa membuatnya konsisten dalam mengidentifikasi sentimen positif. Jumlah besar data yang berhasil diklasifikasikan sebagai sentimen positif sebanyak 753 dari 1000 dataset, menunjukkan ketangguhan SVM. Sebaliknya, *Lexicon Based* mungkin terbatas dalam menangani kompleksitas sentimen positif dan nuansa bahasa yang tidak selalu tercakup dalam daftar kata terlebih dahulu. Akurasi, presisi, dan recall yang tinggi mengonfirmasi kehandalan SVM dalam memahami dan mengklasifikasikan sentimen dalam konteks ini.

## 4. Kesimpulan

Dalam menerapkan metode SVM dan *Lexicon Based* untuk menganalisis sentimen terkait larangan impor pakaian bekas di Indonesia, peneliti memproses dan memberi label pada dataset. Kedua metode diimplementasikan dan dievaluasi pada data uji dengan melibatkan akurasi, presisi, dan recall untuk memahami kinerja klasifikasi sentimen positif dan negatif terhadap kebijakan tersebut. Dari 1000 data tweet yang digunakan, algoritma Support Vector Machine (SVM) mencapai akurasi sebesar 85,87%, dengan recall positif sebesar 93,83% dan recall negatif 62,00%. Kelas precision positif 88,11%, serta class precision negatif 77,02%. Di sisi lain, metode *Lexicon Based* mencapai akurasi 60%, dengan prediksi positif yang akurat sebanyak 518 dan prediksi negatif yang tepat sebanyak 78. Precision untuk kelas negatif adalah 31%, sedangkan untuk kelas positif adalah 69%. Recall kelas negatif sebesar 15%, sedangkan kelas positif mencapai 75%.

Hasil analisis sentimen penerapan metode *Support Vector Machine* untuk membangun model klasifikasi menghasilkan 753 data berhasil diklasifikasikan sebagai sentimen positif, sementara 147 data diklasifikasikan sebagai sentimen negatif. Nilai akurasi yang berbeda menunjukkan bahwa, analisis sentimen menggunakan metode Support Vector Machine

memiliki tingkat akurasi yang lebih tinggi dari pada Lexicon Based. Hal ini menunjukkan bahwa SVM lebih efektif dalam memberikan hasil yang reliabel dalam mengklasifikasikan sentimen.

Dampak bagi pemerintah dari hasil analisis sentimen ini adalah peningkatan pemahaman terhadap pandangan masyarakat terkait larangan impor pakaian bekas. Penggunaan metode *Support Vector Machine* (SVM) menghasilkan tingkat akurasi yang lebih tinggi daripada metode Lexicon Based. Hal ini mengindikasikan bahwa SVM lebih efektif dalam memberikan hasil yang dapat diandalkan dalam mengklasifikasikan sentimen, memungkinkan pemerintah untuk mendapatkan pemahaman yang lebih tepat dan akurat mengenai dukungan atau ketidaksetujuan masyarakat terhadap kebijakan tersebut. Keputusan dan tanggapan yang didasarkan pada analisis sentimen yang lebih akurat dapat membantu pemerintah mengambil langkah-langkah yang lebih tepat dan responsif terhadap dinamika opini masyarakat.

## Daftar Rujukan

- [1] Nasrullah, "Analisis Pemanfaatan Media Sosial," *Pancasila: Jurnal Keindonesiaan*, vol. 02, no. 1, pp. 71–82, 2015.
- [2] D. S. Puspitarini and R. Nuraeni, "Pemanfaatan Media Sosial Sebagai Media Promosi," *Jurnal Common*, vol. 3, no. 1, pp. 71–80, 2019, doi: 10.34010/common.v3i1.1950.
- [3] Telset.id, "10 Aplikasi Media Sosial Terpopuler di Dunia, Paling Banyak Digunakan." 2022.
- [4] N. M. S. Hadna, I. S. Paulus, and W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Seminar Nasional Teknologi Informasi dan Komunikasi*, no. March, pp. 1–8, 2016.
- [5] D. Suparto and A. Habibullah, "Efektivitas Penggunaan Sosial Media Twitter dalam Penyebaran Informasi dalam Pelayanan Publik," *Indonesian Governance Journal : Kajian Politik-Pemerintahan*, vol. 4, no. 2, pp. 161–172, 2021, doi: 10.24905/igj.v4i2.1927.
- [6] F. Zahria Emeraldien, R. Jefri Sunarsono, and R. Alit, "Twitter Sebagai Platform Komunikasi Politik Di Indonesia," *Jurnal Teknologi dan Informasi*, vol. 14, no. 1, pp. 21–30, 2019, [Online]. Available: www.statisticbrain.com
- [7] R. Ariana, "Peraturan Menteri Perdagangan No 40 Tahun 2022," *Permendag*, pp. 1–23, 2022.
- [8] M. A. Sobirin, "Analisis Sentimen Pemberlakuan Karantina di Masa Covid-19 Pada Media Sosial Twitter di Indonesia Menggunakan Algoritme Support Vector Machine (Svm) Dan Lexicon Based," *Nucl. Phys.*, vol. 13, no. 1, pp. 104–116, 2023.
- [9] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [10] F. J. Wahidna and P. Nerisafitra, "Analisis Sentimen Pengguna Sistem Pay Later Menggunakan Support Vector Machine Metode Pembobotan Lexicon," *Journal of Informatics and Computer Science (JINACS)*, vol. 04, pp. 334–343, 2023, doi: 10.26740/jinacs.v4n03.p334-343.
- [11] H. P. P. Zuriel and A. Fahrurrozi, "Implementasi Algoritma Klasifikasi Support Vector Machine Untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan Psbb," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 2, pp. 149–162, 2021, doi: 10.35760/ik.2021.v26i2.4289.
- [12] A. Nuri, "Implementasi Naïve Bayes dan Support Vector Machine dengan Lexicon Based untuk analisis sentimen pada twitter," *Institutional Repository UIN Syarif Hidayatullah Jakarta*, no. 8.5.2017, pp. 2003–2005, 2022.
- [13] D. Gunawan, "Evaluasi Performa Pemecahan Database dengan Metode Klasifikasi Pada Data Preprocessing Data mining," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 2, no. 1, pp. 10–13, 2016, doi: 10.23917/khif.v2i1.1749.
- [14] H. Najjichah, A. Sukur, and H. Subagyo, "Pengaruh Text Preprocessing dan Kombinasinya," *Jurnal Teknologi Informasi*, vol. 15, no. 1, pp. 1–11, 2019, [Online]. Available: [https://download.garuda.kemdikbud.go.id/article.php?article=2887463&val=25402&title=Pengaruh Text Preprocessing dan Kombinasinya Pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia](https://download.garuda.kemdikbud.go.id/article.php?article=2887463&val=25402&title=Pengaruh%20Text%20Preprocessing%20dan%20Kombinasinya%20Pada%20Peringkat%20Dokumen%20Otomatis%20Teks%20Berbahasa%20Indonesia)
- [15] A. Amalia, M. S. Lydia, S. D. Fadilla, and M. Huda, "Perbandingan Metode Klaster dan Preprocessing Untuk Dokumen Berbahasa Indonesia," *Jurnal Rekayasa Elektrika*, vol. 14, no. 1, pp. 35–42, 2018, doi: 10.17529/jre.v14i1.9027.
- [16] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [17] N. L. W. S. R. Ginantra et al., *Machine Learning: Teori dan Metode*. 2023.
- [18] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, p. 1580, 2021, doi: 10.30865/mib.v5i4.3347.
- [19] C. A. Sari, A. Sukmawati, R. P. Aprilli, P. S. Kayaningtias, and N. Yudistira, "Perbandingan Metode Naive Bayes, Support Vector Machine Dan Decision Tree Dalam Klasifikasi Konsumsi Obat," *Jurnal Litbang Edusaintech*, vol. 3, no. 1, pp. 33–41, 2022.
- [20] D. Ramadhan and E. B. Setiawan, "Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," .... *Telkomuniversity.Ac.Id*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>
- [21] G. N. Aulia and E. Patriya, "Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019," *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 2, pp. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.
- [22] M. Al Khadafi, Kurnia Paraniha Kartika, and Filda Febrinita, "Penerapan Metode Naïve Bayes Classifier Dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada Bpjs," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 725–733, 2022, doi: 10.36040/jati.v6i2.5633.
- [23] D. Hernikawati, "Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis," *Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi*, vol. 23, no. 1, pp. 21–31, 2021, [Online]. Available: <http://dx.doi.org/10.33169/iptekkom.23.1.2021.21-31>
- [24] N. S. Fathullah, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naive Bayes dengan Fitur Lexicon-Based," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, pp. 590–593, 2020, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6987>
- [25] L. Hermawan and M. Bellanier Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *Jurnal Transformatika*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.