

Analisis Sentimen Penggunaan Twitter Terhadap Penggunaan Cairan Desinfektan Menggunakan Metode Term Frequency–Inverse Document Frequency Dan Support Vector Machine

¹Hafez Aditya, ²Ardiansyah, ³Sidik, ⁴Windu Gata

^{1,2,3,4} STMIK Nusa Mandiri, Jl. Kramat Raya Nomor 18, RT 05 RW 07, Kwitang, Kecamatan Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450 Indonesia

email : ¹14002325@nusamandiri.ac.id, ²14002404@nusamandiri.ac.id, ³sidik.sdk@nusamandiri.ac.id, ⁴windu@nusamandiri.ac.id

ABSTRACT

Polemic against the pros and cons of the use of disinfectant fluid to break the distribution chain of COVID 19 is currently in a public talk, no exception on social media platforms, especially Twitter. The opinion can be used as a sentiment analysis material for the assessment and the pros and cons of disinfectant fluid to break down the corona virus chain as the cause of the COVID-19 pandemic, whether positive or negative, and knowing what opinion factors often arise. The results of these sentiment analyses can help in the broad community's understanding of the importance of correct use of disinfectant liquids to face the COVID19 outbreak.

Keywords – Desinfektan, SVM, Twitter, COVID-19

ABSTRAK

Polemik terhadap pro dan kontra penggunaan cairan desinfektan untuk memutus persebaran rantai penularan COVID 19 saat ini sedang marak dibicarakan di masyarakat, tak terkecuali juga di platform media sosial, khususnya twitter. Opini tersebut dapat dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui penilaian serta pro dan kontra terhadap cairan desinfektan untuk memutus persebaran mata rantai virus corona sebagai penyebab pandemic COVID-19, apakah positif atau negatif, serta mengetahui faktor opini apa yang sering muncul. Hasil dari analisis sentimen tersebut dapat membantu dalam pemahaman masyarakat luas akan pentingnya penggunaan cairan desinfektan dengan benar untuk menghadapi wabah covid19.

Kata Kunci – Desinfektan, SVM, Twitter, COVID-19

1. Introduction

Pada awal tahun 2020 ini kita semua dikejutkan dengan wabah COVID 19 yang berasal dari virus corona. Cukup mengejutkan karena dampaknya yang begitu luas kepada masyarakat. Karena dampak yang meluas dan penyebaran yang begitu cepat, banyak masyarakat yang pada akhirnya sangat memperdulikan hal tersebut.

Salah satu upaya yang dapat dilakukan oleh warga masyarakat adalah upaya-upaya pencegahan, yaitu dengan cara memperhatikan kebersihan dan kesehatan. Untuk menunjang hal ini, masyarakat banyak memburu cairan desinfektan, yang berfungsi sebagai pembunuh bakteri dan kuman dalam upaya pencegahan tersebut[1].

Pro dan Kontra penggunaan cairan tersebut pada akhirnya bermunculan, karena menurut penelitian ahli kesehatan, menyimpulkan bahwa penggunaan cairan desinfektan dengan cara disemprotkan langsung ke tubuh dan beberapa objek lain dengan tujuan untuk membunuh virus ternyata tidak sepenuhnya aman. Beberapa efek samping akan muncul seperti iritasi kulit, bahaya terhadap pernafasan jika terhirup. Selain memiliki efek samping, penyemprotan cairan desinfektan ternyata juga jika disemprotkan di objek terbuka seperti jalan raya, bangku halte an pagar juga tidak efektif dalam membunuh virus seperti virus corona sebagai penyebab pandemic COVID-19[2][3].

Perkembangan teknologi dewasa ini yang sangat pesat tentunya akan memberikan dampak bagi sendi kehidupan lain, baik itu dampak positif maupun dampak negatif. Teknologi internet merupakan salah satu bidang teknologi yang memiliki perkembangan paling pesat diantara teknologi yang lain. Beberapa sendi kehidupan yang mendapat dampak dari pesatnya perkembangan internet adalah komunikasi [4].

Penggunaan media sosial di Indonesia semakin meningkat dari hari ke hari. Saat ini ada sekitar 132,7 juta orang Indonesia yang aktif di media sosial baik di Facebook, WhatsApp dan Twitter, berdasarkan data dari We Are Social yang dilansir inet.detik.com[5]. Pengaruh cuitan di media sosial dapat menyebar secara cepat. Setiap akun dapat menjadi sumber berita, setiap akun juga bisa menjadi sumber fitnah ataupun penerus fitnah. Khususnya untuk platform Twitter hingga kini masih saja menjadi yang diminati banyak orang, tak terkecuali masyarakat Indonesia.

Pada platform twitter, pengguna bisa membicarakan banyak hal dan tema, baik itu tema serius seperti politik, keamanan, pertahanan negara, sampai hal-hal ringan seperti dunia hiburan, travelling dan kuliner. Termasuk pada saat ramai isu pandemik Covid-19, banyak hal dibicarakan di semua media sosial, tak terkecuali platform twitter ini [6]. Twitter merupakan situs web layanan jejaring sosial yang banyak diminati pengguna internet sebagai media komunikasi dan mendapatkan informasi. Informasi yang terdapat pada Twitter berupa pertanyaan, opini atau komentar, baik yang bersifat positif maupun negative [4] [7].

Pada penelitian sebelumnya dari penelitian yang dilakukan mengenai Analisa Sentimen di Twitter mengenai operasi tangkap tangan Komisi Pemberantasan Korupsi (KPK) masing-masing dataset memiliki 78 data tweet positif dan 78 data tweet negative. Maka dapat disimpulkan untuk data tweet dengan metode Particle Swarm Optimize untuk Support Vector Machine (PSO) menghasilkan akurasi 83.79% dan AUC 0.910 adalah yang paling tinggi dibandingkan algoritma Naïve Bayes, Naïve Bayes (PSO), Support Vector Machine. Dari hasil pengujian data tweet dengan penggunaan metode Support Vector Machine baik yang belum dioptimasi maupun yang sudah dioptimasi dengan Particle Swarm Optimization selalu menghasilkan akurasi yang lebih baik dari metode Naïve Bayes [8].

Pada penelitian lain mengenai Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS, hasil pada akurasi seleksi prediksi YA sejumlah 772 benar dan tidak sesuai sejumlah 32, sehingga presisinya adalah 96.02%. Sedangkan prediksi TIDAK mempunyai hasil sejumlah 0 kesalahan dan 14 benar dalam prediksi TIDAK. Sehingga hasil akurasi yang didapatkan sejumlah 96.15%. Sehingga penggunaan dari aplikasi SMS-Gateway yang digunakan di lingkungan LKBN ANTARA telah sesuai dengan kebutuhan selama bulan November 2015. Serta dapat membantu LKBN ANTARA dalam melakukan klasifikasi isi SMS dengan klasifikasi YA dan TIDAK.[9]

Berbeda dengan penelitian yang telah dilakukan sebelumnya, tujuan dari penelitian ini adalah untuk mengetahui gambaran sentimen positif dan negatif terkait dengan penggunaan cairan desinfektan untuk mencegah memutus mata rantai persebaran virus corona yang menjadi penyebab pandemic COVID-19 dimana tweet diambil untuk mewakili pendapat dari masyarakat

2. Research Method

2.1 Analisis Sentimen

Analisis sentimen atau penambangan pendapat adalah suatu bidang studi untuk menganalisis pendapat orang terhadap suatu entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik. Analisis sentimen ini melihat pada pendapat seseorang atau institusi yang mengemukakan suatu sentimen positif atau negatif. Oleh karena itu, penelitian dalam analisis sentimen tidak hanya memiliki dampak penting pada NLP (Natural Language Processing), tetapi

mingkin juga memiliki dampak mendalam pada ilmu manajemen, politik, ekonomi, dan ilmu sosial karena mereka semua dipengaruhi oleh pendapat orang [6].

2.2 Text Mining

Text mining bertujuan untuk menemukan pola yang tersembunyi pada sumber tertentu agar dapat digunakan untuk suatu tujuan. Untuk melakukan analisis pembuat keputusan, maka dibutuhkan data tidak terstruktur yang berjumlah besar dalam bentuk dokumen. Text mining bukanlah sebuah fungsi, akan tetapi kumpulan dari berbagai macam fungsi yang dikombinasikan dan disebut fungsi text mining. Fungsi utama dari text mining meliputi Searching, Information Extraction, Categorization, Summarization, Prioritization, Clustering, Information Monitor dan Question & answers [10], [11].

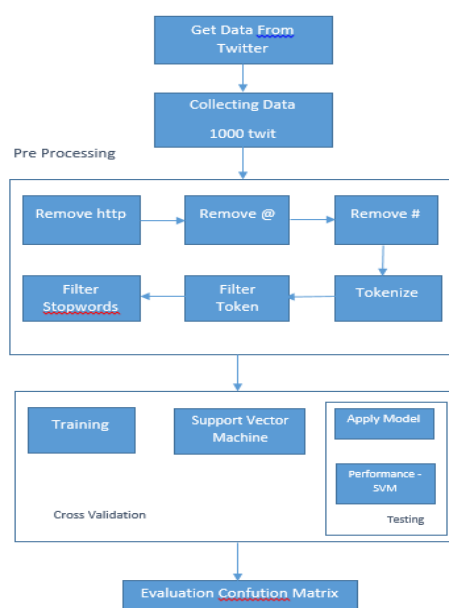
2.3 Support Vector Machine

Metode Support Vector Machine (SVM) adalah metode klasifikasi linier dengan menemukan hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space (Saifinnuha, 2015). Prinsip dasar dari SVM adalah pengklasifikasi linear, kemudian dikembangkan menjadi pengklasifikasi non-linear dengan memasukkan kernel trik pada ruang dimensi tinggi [6], [11], [12].

2.4 Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency)

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) adalah suatu proses untuk melakukan transformasi data dari data teks ke dalam data numerik untuk dilakukan pembobotan pada tiap kata. TF- IDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen. TF adalah frekuensi kemunculan kata pada di tiap 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 inverse dari nilai DF. Hasil dari pembobotan kata menggunakan TF-IDF ini adalah hasil perkalian dari TF dikalikan dengan IDF. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen.[6][14]

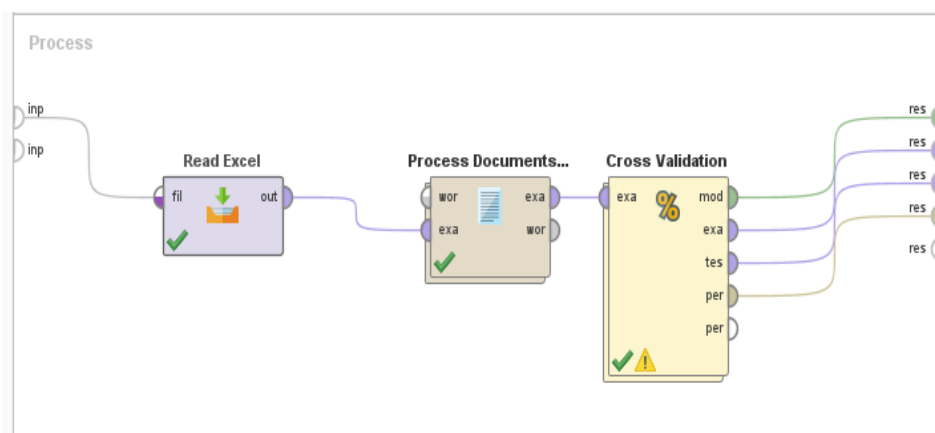
Metode penelitian yang akan digunakan pada penelitian ini sebagaimana ditunjukkan pada gambar 1 dibawah ini :



Gambar 1. Metodologi Penelitian Analisa Sentimen Prokontra Disinfektan

Tahap pertama yang dilakukan adalah dengan mengumpulkan data tweet (data collection). Setelah mendapatkan data tweet, tahap selanjutnya adalah pre-processing agar data siap untuk diproses. Dari dataset yang sudah didapat, selanjutnya data akan diolah menggunakan aplikasi RapidMiner 9.6. Data pertama-tama akan dilakukan pembersihan data terlebih dahulu yang bertujuan untuk menghindari duplikasi, mengeliminasi data dan lain-lain, teknik pembersihan tersebut disebut juga teknik preprocessing.

Berikutnya dilakukan proses cross validation yang meliputi training, support vector machine dan testing (apply model dan performance SVM). Kemudian data akan dilakukan proses pemodelan dengan menggunakan SVM. SVM merupakan sebuah metode pembelajaran mesin yang yang dapat digunakan untuk melakukan klasifikasi teks secara otomatis. Aplikasi SVM untuk pengklasifikasian teks secara otomatis dirintis oleh Thorsten Joachim dengan menggunakan bobot index term sebagai fitur. Untuk desain model pemrosesan data adalah sebagai berikut dijelaskan berikut ini

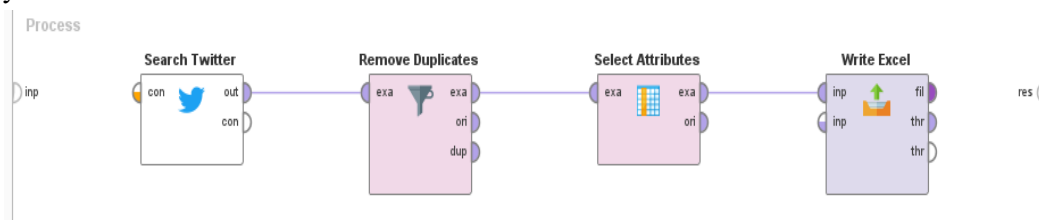


Gambar 2. Desain model SVM

3. Result and Analysis

3.1 Pengumpulan Data

Untuk pengumpulan data pada penelitian ini, diambil dari platform social media yaitu twitter. Dengan menggunakan tools RapidMiner 9.6 untuk mendapatkan data tweet dengan parameter query “desinfektan” dan didapatkan 1000 twit atau cuitan yang membahas dan mengandung kata “desinfektan”. Dari 1000 twit yang dapat diambil, masih perlu dilakukan proses menghilangkan duplikasi dan select attribute, sebagaimana diperlihatkan pada gambar berikut proses pengumpulan datanya



Gambar 3. Alur Pengumpulan Data

3.2 Data Cleansing

Setelah mendapatkan data tweet, langkah pertama dilakukan data cleansing terlebih dahulu dengan tujuan untuk menghilangkan duplikasi yang terjadi. Dari 1000 tweet, terdapat 178 tweet termasuk duplikasi dan dilakukan penghapusan sehingga menyisakan 822 tweet. Proses dilanjutkan dengan menghilangkan “RT “ pada data tweet.

3.3 Preprocessing

Proses dilanjutkan dengan tahap pre-processing menggunakan aplikasi RapidMiner 9.6 dengan parameter vector creation “TF-IDF” agar data siap untuk diproses. Pada tahap pre-processing, dilakukan beberapa proses sebagai berikut :

3.3.1 Proses transform cases (to Lower case) untuk mengubah teks menjadi huruf kecil semua;

Tabel 1. Transform case (to lower case)

No.	Sebelum	Sesudah
1	@WHOIndonesia: #Indonesia, jgn menyemprot disinfektan langsung ke badan seseorang, karena hal ini bisa membahayakan. Gunakan disinfektan...	@whoindonesia: #indonesia, jgn menyemprot disinfektan langsung ke badan seseorang, karena hal ini bisa membahayakan. gunakan disinfektan...

3.3.2 Remove http untuk menghilangkan “http” pada teks

Tabel 2. Remove http

No.	Sebelum	Sesudah
1	@HaiMagazine: WHO Indonesia Larang Bilik Disinfektan Disemprotkan ke Manusia https://t.co/oI92qLGmsI https://t.co/wUIWCGFGqi	@haimagazine: who indonesia larang bilik disinfektan disemprotkan ke manusia

3.3.3. Remove @ untuk menghilangkan “@” pada teks;

Tabel 3. Remove @

No.	Sebelum	Sesudah
1	@kompascom: Dibandingkan cairan disinfektan, sabun bayi dinilai lebih aman dan ampuh untuk tangkal virus corona pada kabin mobil https://...	kompascom: dibandingkan cairan disinfektan, sabun bayi dinilai lebih aman dan ampuh untuk tangkal virus corona pada kabin mobil

3.3.4. Remove Tanda # untuk menghilangkan selain kata (angka, tanda baca,dll) pada teks;

Tabel 4. Remove Tanda #

No.	Sebelum	Sesudah
1	muhammadiyah: Kolaborasi Alfamart dan Lazismu Lakukan Penyemprotan Disinfektan di Rumah Ibadah https://t.co/k38LzmTI0N #Muhammadiyah #...	muhammadiyah: kolaborasi alfamart dan lazismu lakukan penyemprotan disinfektan di rumah ibadah

3.3.5. Filter tokens (by length) dengan parameter *min chars* 4 dan *max chars* 25 untuk membatasi huruf pada kata minimal 4 dan maksimal 25 pada teks;

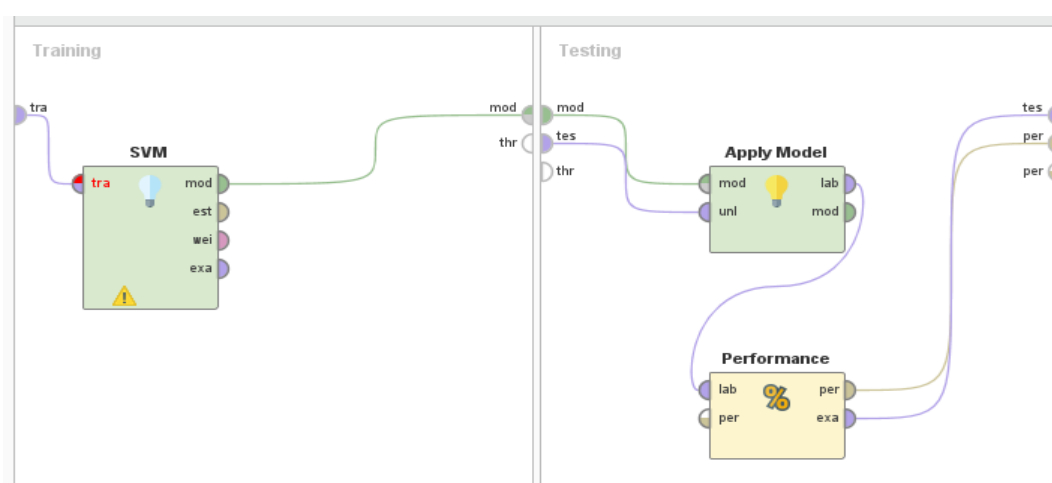
3.3.6. *Filter stopwords (dictionary)* untuk menghilangkan kata-kata yang tidak memiliki makna pada teks.

Tabel 5. *Filter Stopwords*

No.	Sebelum	Sesudah
1	Aku pake cairan disinfektan biar steril	aku cairan disinfektan steril

3.4 Cross Validation

Langkah berikutnya dilakukan adalah proses cross validation dengan menggunakan algoritma SVM dan yang meliputi proses training dan testing (apply model dan performance-SVM). Setelah itu dilakukan proses pemodelan dengan menggunakan Algoritma SVM untuk mendapatkan nilai accuracy, precision dan recall berdasarkan performance vector (performance-SVM).



Gambar 4. Cross Validation SVM

Dari proses cross validation pada tools RapidMiner 9.6, yaitu training dan testing, algoritma SVM menghasilkan nilai accuracy 78.93%, precision 45.83% dan recall 6.41%. Dapat kita lihat bahwa untuk prediksi positif dari 797 jumlah data, 637 dinyatakan true positive dan 160 dinyatakan true negatif dengan tingkat akurasi 79,92%. Sementara total prediksi negatif dari 24 data, 13 dinyatakan true positif dan 11 dinyatakan true negatif dengan tingkat akurasi 45,83%

Criterion: ☒ Table View ☐ Plot View			
accuracy	accuracy: 78.93% +/- 1.14% (micro average: 78.93%)		
precision			
recall			
AUC (optimistic)			
AUC			
AUC (pessimistic)			
	true Positif	true Negatif	class precision
pred. Positif	637	160	79.92%
pred. Negatif	13	11	45.83%
class recall	98.00%	6.43%	

Gambar 5 . Nilai accuracy algoritma SVM

Untuk mengukur tingkat akurasi perlu ditambahkan hasil recall dan precission. Precision merupakan tingkat akurasi antara informasi yang diminta oleh pengguna dan jawaban yang dihasilkan oleh sistem. Hasil precission dapat dilihat pada Gambar 5 Sementara recall adalah tingkat keberhasilan sistem dalam menemukan informasi yang dapat dilihat pada Gambar 6

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Table ViewPlot View

precision: 45.83% (positive class: Negatif)

	true Positif	true Negatif	class precision
pred. Positif	637	160	79.92%
pred. Negatif	13	11	45.83%
class recall	98.00%	6.43%	

Gambar 6 . Nilai *precision* algoritma SVM

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

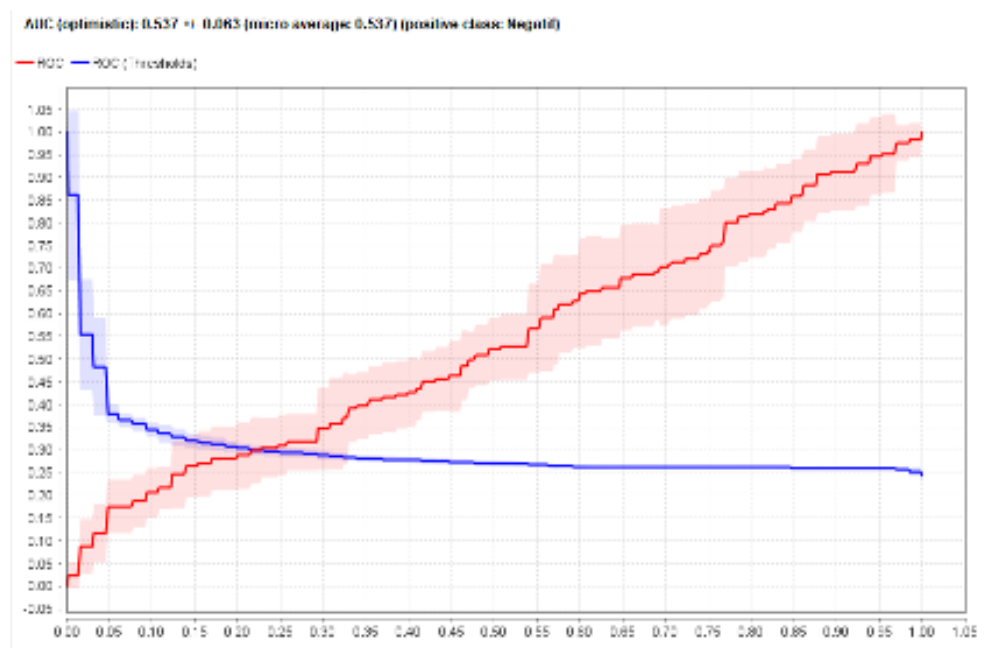
Table ViewPlot View

recall: 6.41% +/- 6.42% (micro average: 6.43%) (positive class: Negatif)

	true Positif	true Negatif	class precision
pred. Positif	637	160	79.92%
pred. Negatif	13	11	45.83%
class recall	98.00%	6.43%	

Gambar 7 . Nilai *recall* algoritma SVM

Pada gambar kurva berikut bisa diketahui untuk cross validation SVM adalah AUC (optimistic): 0.537 +/- 0.063 (micro average: 0.537) (positive class: Negatif).



Gambar 8. AUC Process SVM

Sebanyak 1000 twitt dan setelah dilakukan pre precessing sehingga pada akhirnya yang didapatkan hanya 822 twitt dalam penelitian ini, maka beberapa kesimpulan yang didapatkan adalah sebagai berikut algoritma SVM menghasilkan nilai accuracy 78.93%, precision 45.83% dan recall 6.41%. Dan hasilnya menunjukan bahwa pada masa pandemik corona dan merebaknya COVID-19 saat ini, masyarakat meyakini dan melakukan pencegahan dini dengan bergantung pada cairan desinfektan untuk menekan mata rantai persebaran COVID-19.

4. Conclusion

Berdasarkan analisis dan pengujian yang dilakukan dalam penelitian mengenai Sentimen Analisis Pro Kontra Penggunaan Desinfektan Saat Wabah Covid-19 Melanda Pada Platform Media Sosial Twitter dengan Menggunakan Metode TF-IDF Dan Support Vector Machine dengan mengambil sampling twitter dengan query “desinfektan”. Dapat kita ketahui bahwa dalam masa pandemic COVID 19, masyarakat secara umum masih berpendapat jika penggunaan cairan desinfektan dapat digunakan sebagai salah satu alat untuk memutus mata rantai persebaran virus corona. Walaupun penggunaan cairan desinfektan sebagai alat untuk memutus mata rantai persebaran virus corona belum terbukti secara medis dan klinis, karena penelitian terkait hal ini juga masih belum banyak dilakukan

References

- [1] T. Joko, O. Setiani, and A. Ariani, “Efektivitas Dosis Desinfektan Fenol Terhadap Angka Kuman Pada Lantai Ruang Rawat Inap Rsud Tugurejo Kota Semarang,” *J. Kesehat. Masy. Univ. Diponegoro*, vol. 3, no. 1, p. 18515, 2015.
- [2] A. Adlia and Dkk, “Tanggapan terhadap maraknya penggunaan disinfektan pada bilik disinfeksi untuk pencegahan COVID-19,” pp. 4–7, 2020.
- [3] K. K. RI, 610.28 Ind. 12AD.
- [4] M. N. Fatanti, “Twitter dan Masa Depan Politik Indonesia: Analisis Perkembangan Komunikasi Politik Lokal Melalui Internet Twitter and the Future of Indonesian Politics: Analysis of the Development of Local Political Communication over the Internet,” vol. 16, no. 1, pp. 17–28, 2014.
- [5] A. T. Haryanto, “130 Juta Orang Indonesia Tercatat Aktif di Medsos,” *inet.detik.com*, 2018. [Online]. Available: <https://inet.detik.com/cyberlife/d-3912429/130-juta-orang-indonesia-tercatat-aktif-di-medsos>.
- [6] J. A. Septian, T. M. Fahrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor,” *J. Intell. Syst. Comput.*, no. September, pp. 43–49, 2019.
- [7] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017, doi: 10.1074/jbc.M209498200.
- [8] M. I. Komputer and K. J. Pusat, “Sentimen Analisis Operasi Tangkap Tangan KPK Menurut Masyarakat Menggunakan Algoritma Support Vector Machine , Naive Bayes Berbasis Particle Swarm Optimization,” vol. 12, no. 3, pp. 230–243, 2019, doi: 10.30998/faktorexacta.v12i3.4992.
- [9] W. Gata, “Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS,” vol. 6, pp. 1–13, 2017.
- [10] A. Mustafa, A. Akbar, and A. Sultan, “Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization,” *Int. J. Multimed. Ubiquitous Eng.*, vol. 4, no. 2, pp. 183–188, 2009.
- [11] I. Indriati and A. Ridok, “Sentiment Analysis for Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwkn),” *J. Environmental Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 23–32, 2016, doi: 10.21776/ub.jeest.2016.003.01.4.
- [12] I. M. B. S. Darma, R. S. Perdana, and Indriati, “Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 998–1007, 2018.
- [13] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, “Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2909–2916, 2018.
- [14] R. Melita, V. Amrizal, H. B. Suseno, and T. Dirjam, “(TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web,” vol. 11, no. 2, 2018.