

Analisis Sentimen Masyarakat Kalimantan Tengah Terhadap Perkebunan Kelapa Sawit Menggunakan TF-IDF dan Support Vector Machine

Kurniawan Tri Putra¹, Syadza Anggraini², Linda Sutriani³, Suradji⁴, Ali Impron⁵

^{1,2,3,4,5}

Jurusan Informatika, Fakultas Teknik dan Pertanian, Universitas Muhammadiyah Sampit, Indonesia

*Corresponding Author e-mail: kurniawan_tri_putra@umsa.ac.id, anggrainisyadza@gmail.com,
lindasutriani19@gmail.com, msoadradjie@gmail.com, ali.impron@gmail.com

Abstract: *The transformation towards Society 5.0 has had a significant impact on the rapid growth of data available worldwide, both useful and less directly beneficial, known as big data. This phenomenon provides opportunities for researchers to leverage big data as a valuable source of information, provided it is processed and analyzed using appropriate methods. One of the rapidly growing applications is sentiment analysis, which extracts insights from text data, such as that gathered from social media platforms. This study applies the TF-IDF feature extraction technique and the SVM (Support Vector Machine) classification method to perform sentiment analysis on Twitter text data. The results of the research show that the model built using the combination of TF-IDF and SVM achieved an accuracy of 86%, with precision, recall, and F1-Score values of 85% each. These findings indicate that the application of TF-IDF with SVM provides optimal performance in sentiment analysis, considering the word frequency within documents, and makes a significant contribution to processing big data for more accurate and effective sentiment analysis*

Key Words: Big Data, Sentiment Analysis, TF-IDF, SVM

Abstrak: Transformasi menuju Society 5.0 memberikan dampak yang signifikan terhadap pesatnya pertumbuhan data yang tersedia di seluruh dunia, baik yang berguna maupun yang kurang bermanfaat secara langsung, yang dikenal dengan istilah big data. Fenomena ini memberikan peluang bagi peneliti untuk memanfaatkan big data sebagai sumber informasi yang berharga, asalkan diolah dan dianalisis menggunakan metode yang tepat. Salah satu aplikasi yang berkembang pesat adalah analisis sentimen, yang mengekstraksi wawasan dari data teks, seperti yang dikumpulkan dari platform media sosial. Penelitian ini menerapkan teknik ekstraksi fitur TF-IDF dan metode klasifikasi SVM (Support Vector Machine) untuk melakukan analisis sentimen pada data teks Twitter. Hasil penelitian menunjukkan bahwa model yang dibangun menggunakan kombinasi TF-IDF dan SVM mencapai akurasi sebesar 86%, dengan nilai presisi, recall, dan F1-Score masing-masing sebesar 85%. Temuan ini menunjukkan bahwa penerapan TF-IDF dengan SVM memberikan kinerja yang optimal dalam analisis sentimen, mempertimbangkan frekuensi kata dalam dokumen, dan memberikan kontribusi signifikan dalam pemrosesan data besar untuk analisis sentimen yang lebih akurat dan efektif

Kata Kunci: Big Data, Analisis Sentimen, TF-IDF, SVM

Pendahuluan

Seiring dengan perkembangan zaman yang semakin pesat, konsep *Society 5.0* kini menjadi nyata, di mana teknologi modern dan ilmu pengetahuan digunakan oleh seluruh umat manusia untuk memenuhi kebutuhan serta mempermudah berbagai aktivitas. Salah satu dampak besar dari perubahan ini terlihat dalam bidang *big data*, yang semakin penting mengingat banyaknya pengguna media sosial yang kini memanfaatkan platform-platform tersebut untuk berbagai keperluan, mulai dari berkomunikasi, berjualan, berbagi informasi, hingga sebagai saluran untuk mengekspresikan perasaan atau berbagi cerita. Adanya sumber *big data* ini memberikan peluang bagi penelitian untuk memanfaatkan informasi yang terkumpul, dengan salah satunya menggunakan metode *machine learning* untuk melakukan analisis sentimen yang dapat menghasilkan wawasan lebih berguna.

Analisis sentimen dalam pemanfaatan *big data* dilakukan dengan mengumpulkan dan mengolah data teks yang berasal dari pengguna media sosial, yang mencerminkan perasaan atau ekspresi mereka terhadap suatu objek atau topik yang sedang dibahas. Berdasarkan penelitian (Arfat et al., 2022), salah satu platform media sosial yang paling banyak digunakan oleh masyarakat Indonesia adalah Twitter, dengan persentase pengguna mencapai 56%, setelah mengoptimalkan kinerja sistem analisis sentimen menggunakan *machine learning*, dengan

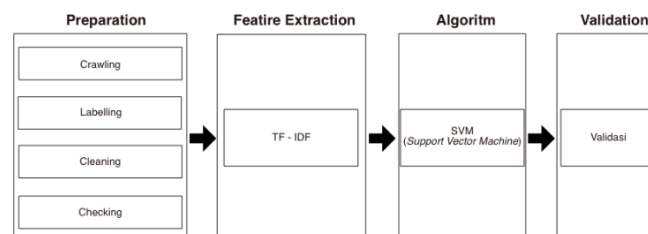


YouTube, WhatsApp, Facebook, dan Instagram. Penelitian yang mengarah pada analisis sentimen dengan data yang bersumber dari Twitter kini sedang berkembang pesat dan telah banyak diimplementasikan di Indonesia (Romadhoni & Holle, 2022) (Soemedhy et al., 2022) (Putri & Fatonah, 2022) (Setyawinda et al., 2020). Penelitian-penelitian ini bertujuan untuk fokus pada penggabungan teknik dan metode yang efektif. Namun, seringkali penelitian tersebut hanya mengutamakan metode yang terbaik tanpa mempertimbangkan faktor lain yang juga berpengaruh terhadap hasil kinerja *machine learning*, seperti pada proses ekstraksi fitur.

Berdasarkan permasalahan yang telah dibahas, salah satu teknik yang umum digunakan dalam ekstraksi fitur adalah TF-IDF (Gifari et al., 2022). Pada metode ini memiliki peran penting dalam merepresentasikan data teks dalam bentuk numerik, namun masih tetap memiliki kelebihan dan kekurangan. Penelitian ini akan membahas dengan penggabungan teknik tersebut yang dipadukan dengan metode *Support Vector Machine* (SVM) untuk mengoptimalkan kinerja *machine learning* dalam menganalisis sentimen. Data yang digunakan berasal dari media sosial Twitter, dengan fokus pada studi kasus Analisis Sentimen Masyarakat.

Metode Penelitian

Penelitian tentang analisis sentimen menggunakan *machine learning* ini dirancang dengan melalui beberapa tahap, yaitu dimulai dari pengumpulan data, pelabelan data, pra-pemrosesan data, ekstraksi fitur data, klasifikasi data, evaluasi dan pengujian, kemudian diakhiri dengan perbandingan hasil pengujian dan penarikan Kesimpulan (Septian et al., 2019).



Gambar 1. Alur penelitian

Diagram alir di atas menggambarkan serangkaian langkah yang diperlukan untuk melaksanakan proses analisis sentimen menggunakan *machine learning*, dengan penjelasan sebagai berikut:

Preparation

Pada tahap ini, dilakukan pengumpulan data mentah dari Twitter yang kemudian diproses menjadi data yang siap untuk dianalisis oleh sistem, melalui beberapa langkah yang harus dilalui terlebih dahulu:

1. Crawling

Proses pengambilan data atau penambahan data dilakukan dengan mengakses aplikasi media sosial Twitter menggunakan Google Colab dan bahasa pemrograman Python. Data yang diambil berupa teks dari tweet pengguna yang dikelompokkan berdasarkan topik yang relevan, yaitu Perkebunan Kleapa Sawit, yang difokuskan pada hashtag (*#PerkebunanKelapaSawitKalimantanTengah*), Dengan data Sebanyak 300 tweet dikumpulkan untuk dianalisis.

2. Labelling

Proses pelabelan dilakukan pada data yang telah dikumpulkan dari Twitter, yang berjumlah 300 tweet. Pada tahap ini, pelabelan dilakukan secara manual oleh pihak yang memiliki keahlian di bidang hukum dan perundang-undangan. Data diberi label berdasarkan dua kategori, yaitu label negatif untuk tweet yang mengandung makna atau konotasi yang buruk, dan label positif untuk tweet yang memiliki makna baik.

3. Cleaning

Pada tahap ini, dilakukan pembersihan data dengan menghapus karakter-karakter yang tidak relevan dan tidak memberikan kontribusi terhadap analisis sentimen. Beberapa langkah yang dilakukan meliputi case folding, yaitu mengubah semua teks menjadi huruf kecil; *remove punctuation*, yang mencakup penghapusan URL, karakter khusus, angka, dan sebagainya; *stopword removal*, yaitu penghapusan kata-kata hubung serta *stemming*, yaitu proses untuk menghilangkan imbuhan.

4. Data checking

Tahap terakhir dalam persiapan data adalah pemeriksaan data, yang bertujuan untuk memastikan bahwa setelah proses pelabelan dan pembersihan, tidak terdapat data yang kosong atau blank, data duplikat, atau data yang tidak seimbang (*imbalance*).

Feature Extraction

Pada tahap ini, dilakukan ekstraksi kata yang diubah menjadi angka dalam bentuk vektor, karena komputer hanya dapat memproses dan mengenali data dalam bentuk numerik, tetapi angka-angka tersebut tetap merepresentasikan makna.

TF-IDF (*Term Frequency - Inverse Document Frequency*) adalah suatu metode untuk menghitung atau mengekstraksi kata-kata menjadi nilai numerik dalam bentuk vektor yang digunakan untuk mengukur pentingnya suatu kata dalam sebuah dokumen atau kumpulan dokumen (korpus). Nilai ini menunjukkan sejauh mana relevansi kata tersebut terhadap dokumen yang dianalisis (Yutika et al., 2021) (Styawati et al., 2021). Pada dasarnya, perhitungan TF-IDF terdiri dari dua komponen utama, yaitu TF (Term Frequency) dan IDF (Inverse Document Frequency), yang masing-masing memiliki rumus dan mekanisme perhitungan tersendiri. Hasil akhirnya merupakan gabungan dari nilai TF dan IDF yang dihitung secara terpisah seperti berikut:

a. TF (*Term Frequency*)

Dengan cara kerja menghitung frekuensi jumlah kemunculan kata pada sebuah dokumen. Dalam setiap dokumen memiliki panjang yang berbeda-beda maka nilai TF akan dibagi dengan panjang dokumen:

$$tf_{t,d} = \frac{n_{t,d}}{\text{(Total number of term in document)}} \quad (1)$$

Keterangan

Tf = frekuensi kemunculan kata pada sebuah dokumen

b. IDF (*Inverse Document Frequency*)

Setelah berhasil menghitung nilai TF kita akan menghitung nilai IDF yang merupakan nilai untuk mengukur seberapa penting sebuah kata, dengan beracuan semakin kecil nilai IDF maka akan dianggap semakin tidak penting kata tersebut, begitupun sebaliknya:

$$idf_d = \log \frac{\text{Number of document}}{\text{(Total number of term in document)}} \quad (2)$$

Keterangan

Idf = mengukur penting/tidak sebuah kata dalam dokumen

c. TF-IDF (*Term Frequency - Inverse Document Frequency*)

Setelah mendapatkan nilai TF dan IDF selanjutnya akan menghitung nilai TF-IDF dengan mengalikannya:

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

Keterangan

TF-IDF = hasil penggabungan antara TF dan IDF

SVM (*Support Vector Machine*)

Merupakan teknik pembelajaran yang digunakan oleh mesin untuk melakukan klasifikasi dan regresi. Cara kerjanya adalah dengan memisahkan dua kelas dengan cara memaksimalkan margin atau jarak antara kelas-kelas tersebut. Persamaan dasar dari metode SVM untuk proses ini dapat dijelaskan melalui rumus berikut (Zhafira et al., 2021).

$$y = f(x) = \text{sign}(w \cdot x + b) \quad (c=1 \quad g=1 \quad \text{karnel}=\text{linear}) \quad (4)$$

Dimana x adalah vektor atau vitur dari sebuah data, w adalah vektor bobot yang mengatur arah dan jarak garis atau *hyperplane*, b adalah bias, sedangkan y adalah label kelas prediksi (-1 atau +1), dan sign adalah fungsi dari signum yang menghasilkan -1 tau +1 tergantung kepada apakah $f(x) > 0$ atau $f(x) < 0$.

Validation

Langkah terakhir dalam proses klasifikasi analisis sentimen menggunakan data teks dari Twitter adalah validasi, yang bertujuan untuk menguji dan menilai keberhasilan teknik serta metode yang telah diterapkan. Untuk evaluasi model ini, digunakan teknik *confusion matrix* yang mengacu pada pengukuran seperti akurasi, presisi, *recall*, dan *F1-score*, dengan acuan tabel sebagai berikut (Normawati & Prayogi, 2021).

Tabel I. *confusion matrix*

TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Keterangan:

- a. TP yaitu jumlah data positif yang terklasifikasi benar
- b. TN yaitu jumlah data negatif yang terklasifikasi benar
- c. FN yaitu jumlah data negatif yang terklasifikasi salah
- d. FP yaitu jumlah data positif yang terklasifikasi salah

Accuracy (A) prediksi benar dari *true positif* dan *true negatif*.

$$A = \frac{(TP+TN)}{(TP+FP+FN+TN)} = 100\% \quad (5)$$

Hasil dan Pembahasan

Persiapan Data

Proses mempersiapkan data mulai dari penambangan data twitter hingga mengolahnya, agar siap untuk dilakukan pemrosesan menggunakan *machine learning*.

1. Crawling twitter

Penambangan data twitter yang telah dilakukan didapat hasil 300 tweet dengan pengelompokan hastag *#PerkebunanKelapaSawitKalimantanTengah* pada priode rentan waktu mulai 01 Oktober 2024 sampai dengan 25 Januari 2025, yang berisikan variabel nama akun *account name*, isi teks unggahan *tweet* dan waktu unggah *date* dengan hasil pada tabel berikut:

Tabel III. Hasil penambangan teks tweet

No	Account Name	Tweet	Date
1	@angengen	Membuka lapangan Kerja!!!	01/10/2023 14:45
2	@maskukur	Alhamdulillah Gajihan >>>> 🍌	01/10/2023 15:58
...
300	@dadar_jagung	Jalan banyak rusak, oleh CPO 🚧	13/01/2025 07:15

2. Labelling

Dari hasil penambangan data twitter yang telah dilakukan dengan jumlah data 300 tweet maka akan dilakukan pelabelan secara manual oleh orang yang berkompeten bidang hukum dengan hasil sebagai berikut:

Tabel IV. Hasil labelling teks tweet

No	Tweet	L1	L2	L3	Keputusan
1	Membuka lapangan Kerja!!!	-	-	-	Positif
2	Alhamdulillah Gajihan >>>> 🍌	+	+	-	Positif
...
300	Jalan banyak rusak, oleh CPO 🚧	-	-	-	Negatif

Keterangan:

Pelabelan dilakukan oleh 3 orang dengan kode L1,L2 dan L3 dan keputusan label didapatkan jika jumlah label tidak kurang dari 2.

3. Cleaning

Noise dan fitur yang tidak relevan dalam data teks dapat memengaruhi kinerja pemrosesan. Pada tahap ini, data akan melalui proses pembersihan atau penyederhanaan kata menjadi bentuk baku, yang mencakup beberapa langkah seperti case folding (mengubah semua kata menjadi huruf kecil), remove punctuation (menghapus URL, karakter, angka, dan tanda baca lainnya dari setiap kalimat),

stopword removal (menghilangkan kata hubung yang tidak penting), serta stemming (menghapus imbuhan pada kata). Hasil dari proses ini dapat dilihat pada tabel berikut.

Tabel V. Hasil cleanning teks tweet

No	Tweet	Cleanning
1	Membuka lapangan Kerja!!!	Membuka lapangan Kerja
2	Alhamdulillah Gajihan >>>> 🍌	Alhamdulillah Gajihan
...
300	Jalan banyak rusak, oleh CPO 🛣️ 🛣️	Jalan banyak rusak, CPO

4. Data Checking

Data yang sudah siap untuk diproses akan dilakukan pengecekan lagi terlebih dahulu, agar data benar-benar valid. Pengecekan ini memiliki hasil yang baik dengan tidak ada data kosong dan duplikat serta jumlah antar kedua label positif dan negative yang seimbang, sebagai berikut:

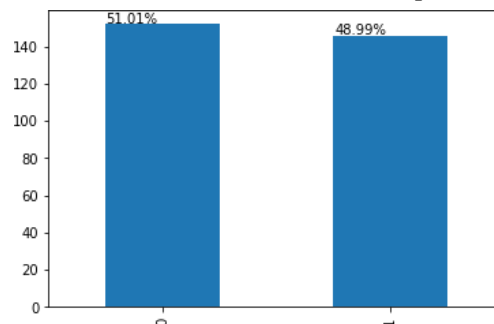
```

df.dropna(inplace=True)
df['tweet'].isna()
0    False
1    False
2    False
3    False
4    False
...
295   False
296   False
297   False
298   False
299   False
Name: tweet, Length: 300, dtype: bool

df.drop_duplicates(inplace=True)
df.duplicated()
0    False
1    False
2    False
3    False
4    False
...
295   False
296   False
297   False
298   False
299   False
Length: 298, dtype: bool

```

Gambar 2. Hasil cek *blank data* dan *duplicated data*



Gambar 3. Hasil cek *imbalance data*

Feature Extraction

Proses perubahan dari data berbentuk teks menjadi sebuah vektor berupa angka untuk dapat diproses oleh computer dengan python dan dibuktikan perhitungan manual (Munir et al., 2018).

1. TF-IDF (*Term Frequency- Inverse Document Frequency*)

Perhitungan atau rumus TF-IDF terbagi menjadi dua yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*) dengan rumus dan cara kerja yang berbeda dan akan digabungkan di akhir perhitungan antara TF dan IDF menjadi TF-IDF (*Term Frequency- Inverse Document Frequency*) sebagai berikut:

DATA UTAMA			
Dokumen 1(D1)	Membuka Lapangan Kerja	Panjang Dokumen	3
Dokumen 2(D2)	Alhamdulillah Gajihan	Panjang Dokumen	2
Dokumen 3(D3)	Jalan Banyak Rusank CPO	Panjang Dokumen	4
Total Dokumen			3

Gambar 4. Sempel data perhitungan manual

- TF (*Term Frequency*)

Hasil perhitungan TF dengan menetapkan rumus yang sudah ada pada sub bab sebelumnya:

Frekuensi kemunculan kata			df	TF		
D1	D2	D3		TF1	TF2	TF3
0	1	0	1	0	0,5	0
0	0	1	1	0	0	0,25
1	0	0	1	0,333333	0	0
1	0	0	1	0,333333	0	0
1	0	1	2	0,333333	0	0,25
0	0	1	1	0	0	0,25
0	0	1	1	0	0	0,25
0	1	0	1	0	0,5	0

Gambar 5. Hasil perhitungan *Term Frequency*

Keterangan:

D1,D2 dan D3 merupakan jumlah frekuensi kemunculan kata pada setiap dokumen dengan hasil df, sedangkan TF1,TF2 dan TF3 merupakan notasi untuk setiap review TF

- IDF (*Inverse Document Frequency*)

Setelah berhasil menghitung nilai TF selanjutnya kita menghitung nilai IDF dari hasil nilai TF:

Frekuensi kemunculan kata			df	TF			IDF
D1	D2	D3		TF1	TF2	TF3	
0	1	0	1	0	0,5	0	0,477121
0	0	1	1	0	0	0,25	0,477121
1	0	0	1	0,333333	0	0	0,477121
1	0	0	1	0,333333	0	0	0,477121
1	0	1	2	0,333333	0	0,25	0,176091
0	0	1	1	0	0	0,25	0,477121
0	0	1	1	0	0	0,25	0,477121
0	1	0	1	0	0,5	0	0,477121

Gambar 6. Hasil perhitungan *Inverse Document Frequency*

Keterangan:

Nilai IDF adalah logaritma dari pembagian jumlah dokumen dengan nilai dari TF.

- TF-IDF (*Term Frequency- Inverse Document Frequency*)

Setelah kedua nilai TF dan IDF didapatkan kemudian akan digabungkan menjadi nilai TF-IDF dengan menyusun angka -angka dengan susunan berupa vektor berikut:

Frekuensi kemunculan kata			df	TF			IDF	TF-IDF		
D1	D2	D3		TF1	TF2	TF3		TF-IDF1	TF-IDF2	TF-IDF3
0	1	0	1	0	0,5	0	0,477121	0	0,238561	0
0	0	1	1	0	0	0,25	0,477121	0	0	0,11928
1	0	0	1	0,333333	0	0	0,477121	0,15904	0	0
1	0	0	1	0,333333	0	0	0,477121	0,15904	0	0
1	0	1	2	0,333333	0	0,25	0,176091	0,058697	0	0,044023
0	0	1	1	0	0	0,25	0,477121	0	0	0,11928
0	0	1	1	0	0	0,25	0,477121	0	0	0,11928
0	1	0	1	0	0,5	0	0,477121	0	0,238561	0

Untuk penerapan dengan menggunakan machine learning berbasis *python* adalah sebagai berikut:

```

✓ [3] from sklearn.feature_extraction.text import TfidfVectorizer
2s

✓ [4] tfidf = TfidfVectorizer()
      respons = tfidf.fit_transform(corpus)
      print (respons)

✓ [5] tfidf.get_feature_names()

✓ [6] respons.todense()
0s
matrix([[0.         , 0.         , 0.62276601, 0.62276601, 0.4736296 ,
         0.         , 0.         , 0.         ],
        [0.70710678, 0.         , 0.         , 0.         , 0.         ,
         0.         , 0.         , 0.70710678],
        [0.         , 0.52863461, 0.         , 0.         , 0.40204024,
         0.52863461, 0.52863461, 0.         ]])

```

Gambar 9. Hasil vektorisasi menggunakan python

Algorithm SVM (*Support Vector Machine*)

Setelah mendapatkan hasil vektor dari data yang sudah melewati beberapa tahapan, pada proses SVM dengan cara kerja memisahkan dua kelas dengan memaksimalkan margin (Wibawa et al., 2018).

Pada proses metode SVM ini dilakukan untuk setiap percobaan feature extraction TF-IDF maupun BOW menggunakan konfigurasi standart sebagai berikut:

```

[ ] from sklearn.svm import SVC

[ ] model1 = SVC(C=1, gamma=1, kernel='linear')
      model1.fit(X_train, y_train)

SVC(C=1, gamma=1, kernel='linear')

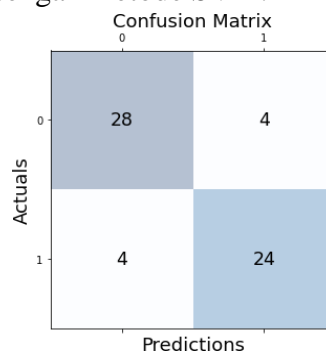
```

Gambar 13. Metode SVM di python

Validation

Proses menguji dan mengukur keberhasilan dari teknik dan metode yang telah diterapkan untuk model percobaan yaitu antara TF-IDF dengan metode SVM (*Support Vector Machine*):

- Berikut hasil validasi TF-IDF dengan metode SVM:



Gambar 14. *Confusion matrix* TF-IDF

Keterangan hasil:

- Berjumlah 28 data positif yang terklasifikasi benar
- Berjumlah 24 data negatif yang terklasifikasi benar
- Berjumlah 4 data negatif yang terklasifikasi salah
- Berjumlah 4 data positif yang terklasifikasi salah

Hasil dari *confusion matrix* akan digunakan sebagai acuan dalam mencari *Accuracy*, *Precision*, *Recall* dan *F1-score*:

Accuracy (0.86) prediksi benar dari *TP* dan *TN*.

Precision (0.85) nilai *true positif* dari seluruh nilai *positif*.

Recall (0.85) persentase prediksi positif dengan *true positif*.
F1-Score (0.85) perbandingan rata-rata *precision* dan *recall*.

Kesimpulan

Kesimpulan dari hasil penelitian ini menunjukkan bahwa penerapan teknik ekstraksi fitur TF-IDF dengan metode SVM (Support Vector Machine) menghasilkan kinerja yang sangat baik dalam analisis sentimen menggunakan data teks Twitter. Model yang dibangun mampu mencapai tingkat akurasi sebesar 86%, dengan nilai *precision*, *recall*, dan *F1-Score* masing-masing sebesar 85%. Hasil ini mengindikasikan bahwa metode yang digunakan, yang mempertimbangkan frekuensi kemunculan kata dalam dokumen, dapat memberikan performa yang optimal. Secara keseluruhan, penerapan teknik TF-IDF dan SVM menunjukkan keunggulan yang signifikan pada setiap pengukuran evaluasi, baik dari sisi akurasi, *precision*, *recall*, maupun *F1-Score*, sehingga menjadi pilihan yang efektif untuk analisis sentimen berbasis teks.

Referensi

- Arfat, M. F., Styawati, S., Nurkholis, A., & Kurniawan, I. (2022). Analisis Sentimen Masyarakat Indonesia Terkait Vaksin Covid-19 Pada Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM). *Jurnal Informatika: Jurnal Pengembangan IT*, 7(2), 96–103.
- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
- Munir, M. M., Fauzi, M. A., & Perdana, R. S. (2018). Implementasi metode backpropagation neural network berbasis lexicon based features dan bag of words untuk identifikasi ujaran kebencian pada Twitter. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3182–3191.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Putri, R., & Fatonah, N. (2022). Perbandingan Metode Klasifikasi serta Analisis Faktor Akademis Pola Kelulusan Mahasiswa di Perguruan Tinggi. *Jurnal Informatika: Jurnal Pengembangan IT*, 7, 109–117. <https://doi.org/10.30591/jpit.v7i2.3082>
- Romadhoni, Y., & Holle, K. F. H. (2022). Analisis Sentimen Terhadap PERMENDIKBUD No. 30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM. *Jurnal Informatika: Jurnal Pengembangan IT*, 7(2), 118–124.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Journal of Intelligent Systems and Computation 43. *J. Intell. Syst. Comput*, 43–49.
- Setyawinda, A. A., Setiyadi, B., & Hartanto, A. D. (2020). Perbandingan Algoritma Word Matching dan Naive Bayes untuk Klasifikasi Sentimen Analisis Komentar Instagram. *Jurnal Informatika: Jurnal Pengembangan IT*, 5(1), 12–16.
- Soemedhy, C. A. A., Trivetisia, N., Winanti, N. A., Martiyaningsih, D. P., Utami, T. W., & Sudio, S. (2022). Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube “Kekerasan Seksual”). *Jurnal Informatika: Jurnal Pengembangan IT*, 7(2), 80–84.
- Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode support vector machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155.
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).
- Yutika, C. H., Adiwijaya, A., & Al Faraby, S. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422–430.
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube.