

Analisis Penerapan Text Mining dan TF-IDF dalam Mengetahui Sentimen Masyarakat Terhadap Kinerja POLRI

Rivalri Kristianto Hondro

¹Fakultas Ilmu Komputer & Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: rivalryhondro@gmail.com

Email Penulis Korespondensi: rivalryhondro@gmail.com
Submitted: 29/04/2023; Accepted: 15/05/2023; Published: 31/05/2023

Abstrak— Analisis sentimen adalah proses menganalisa suatu opini masyarakat terhadap masalah yang ada, dimana isi opini tersebut bisa mengandung sentimen positif atau negatif. Penelitian ini menggunakan data dari suatu platform media sosial yaitu twitter. Opini masyarakat yang terdapat di tweet dimanfaatkan sebagai bahan analisis sentimen masyarakat terhadap kinerja polri. Untuk pengklasifikasian sentimen masyarakat yang mengandung positif atau negatif, maka penulis menggunakan algoritma text mining dan TF-IDF.

Kata Kunci: Polri; kinerja; preprocessing; tf-idf

Abstract— Sentiment analysis is the process of analyzing a public opinion on an existing problem, where the contents of the opinion can contain positive or negative sentiments. This study uses data from a social media platform, namely Twitter. Public opinion contained in the tweet is used as material for analyzing public sentiment on the performance of the National Police. To classify public sentiment as positive or negative, the authors use text mining and TF-IDF algorithms.

Keywords: Polri; performance; preprocessing; tf-idf

1. PENDAHULUAN

Kepolisian Negara Republik Indonesia adalah Polisi Nasional di Indonesia, yang bertanggung jawab langsung di bawah Presiden. Sebelumnya organisasi ini bernama Badan Kepolisian Negara, Djawatan Kepolisian Negara dan Angkatan Kepolisian Republik Indonesia. Dalam pengertian lain POLRI merupakan instrumen negara yang berperan dalam memelihara keamanan dan ketertiban masyarakat, menegakkan hukum serta memberikan perlindungan, pengayoman dan pelayanan masyarakat. Adapun tujuan dibentuknya kepolisian menurut Undang-Undang Dasar Negara 1945 pasal 4 UU No.2 Tahun 2002 adalah untuk mewujudkan keamanan dan ketertiban masyarakat, tertib dan tegaknya hukum, terselenggaranya perlindungan, pengayoman dan pelayanan masyarakat, serta terbinanya ketentraman masyarakat dengan menjunjung tinggi Hak Asasi Manusia [1]. Terkait tugas dan tanggung jawab yang diemban oleh kepolisian, kinerja kepolisian merupakan bagian yang mendapat perhatian dari masyarakat.

Masyarakat telah mengalami perubahan pola pikir yang makin kritis dalam merespon kondisi yang ada, dimana dengan kondisi tersebut maka kinerja polisi dituntut untuk dapat memenuhi kebutuhan masyarakat dalam segala aspek. Masyarakat ingin memberikan beberapa keluhan seperti opini tentang kinerja polisi yang tidak sesuai keinginan masyarakat. Dalam contoh kasus, tagar #PercumaLaporPolisi menjadi trending di twitter karena mencuatnya berita kasus pemerkosaan anak di Luwu Timur pada 6 Oktober 2021 lalu. Kasus tersebut sempat dihentikan oleh polisi. Seiring dengan hal tersebut, masyarakat menaikan tagar #PercumaLaporPolisi sebagai bentuk kekecewaan terhadap polisi dalam menangani kasus tersebut (news.detik.com). Dengan adanya permasalahan tersebut, menimbulkan beberapa opini atau pendapat dari masyarakat terhadap kinerja polisi. Opini yang diberikan oleh masyarakat inilah yang dapat digunakan untuk bahan evaluasi oleh polri untuk meningkatkan kualitas kinerja yang baik kedepannya.

Natural Language Processing (NLP) merupakan proses pembuatan model komputasi dari bahasa sehingga memungkinkan terjadinya interaksi manusia dengan komputer dengan perantaraan alami yang dipakai oleh manusia [2]. Text Mining adalah proses ekstrasi pola (informasi dan pengetahuan) dari sejumlah data yang tidak terstruktur. Terdapat dua masukan untuk text mining yaitu data yang tidak terstruktur (dokumen Word, PDF, kutipan teks) dan data yang terstruktur [3]. Algoritma Term Frequency – Inverse Document Frequency (TF-IDF) merupakan algoritma yang berasal dari bidang information retrieval, namun saat ini semakin banyak digunakan dalam perbandingan dokumen [4].

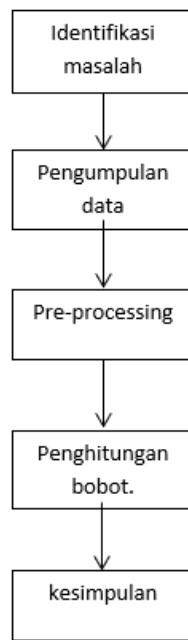
Penelitian terdahulu terkait sentimen telah banyak dilakukan sebelumnya untuk berbagai keperluan. Dalam penelitian Sri Wulan Utami Vitandy, dkk pada jurnal pengembangan teknologi informasi dan Ilmu Komputer vol.3 No.6 tahun 2019 berjudul “Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naive Bayes Classifier” bahwa hasil yang didapatkan dari penelitian tersebut berupa dashboard yang berisi pengelompokan dosen dan klasifikasi metode ini memiliki tingkat presisi 80,3% , penarikan 80,3% dan akurasi 80,1%. Selanjutnya dalam penelitian Ilhamuddin Harahap pada jurnal INTI Vol.9 No.2 tahun 2022 yang berjudul “Analisa Sentimen Masyarakat Terhadap Kondisi New Normal Pasca Pembatasan Sosial Berskala

Besar Akibat Covid 19 Menerapkan Metode Term Frequency-Relevan Frequency” bahwa penerapan metode pada penelitian tersebut cukup memberi informasi tentang analisa masyarakat terhadap kondisi New Normal.

Berdasarkan pada latar belakang tersebut, maka penulis mencoba melakukan penelitian yaitu analisis sentimen masyarakat terhadap kinerja Polri dengan menggunakan algoritma text mining dan TF-IDF untuk mengetahui seberapa besar persentase sentimen positif dan negatif.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Tahapan penelitian pada gambar 1 diatas dapat dijelaskan sebagai berikut:

1. Identifikasi Masalah

Identifikasi masalah yang terkait dengan proses analisis permasalahan yang ada. Permaalahan yang dianalisis yaitu bagaimana menganalisa sentimen masyarakat menggunakan metode tf-idf dan text mining.

2. Pengumpulan Data

Pengumpulan data yang dilakukan dengan mencari komentar masyarakat melalui sebuah platform media sosial yaitu twitter.

3. Pre-processing

Pre-processing yaitu proses menggunakan algoritma text mining memiliki 4 tahapan yaitu case folding, cleaning, tokenizing, stopword, stemming.

4. Perhitungan Bobot

Perhitungan bobot yang terkait proses menghitung bobot kata menggunakan metode TF-IDF.

5. Kesimpulan

2.2 Text Mining

Text Mining adalah bidang penelitian ilmu komputer yang baru dan menarik yang mencoba memecahkan krisis informasi yang berlebihan dengan menggabungkan teknik dari data mining, pembelajaran mesin, pemrosesan bahasa alami, pencarian informasi, dan pengetahuan pengelolaan. Text mining dapat didefinisikan secara luas sebagai proses intensif pengetahuan di mana pengguna berinteraksi dengan dokumen dari waktu ke waktu menggunakan seperangkat alat analisis [5].

2.3 Term Frequency

Term Frequency (TF) adalah bobot suatu kata(t) dalam suatu dokumen(d) dan disimbolkan dengan $tf_{t,d}$. Pendekatan yang paling sederhana dari konsep ini adalah dengan menyatakan bobot suatu kata(t) sebagai jumlah kemunculannya pada dokumen(d). Sebagai contoh, jika dalam suatu dokumen, kata plagiat muncul sebanyak 5 kali maka nilai TF adalah 5.

$$Tf_{t,d} = \frac{\log f_{t,d} + 1}{0} \quad (1)$$

Keterangan:

$$Tf_{t,d} = \text{term frequency}$$

$$F_{t,d} = \text{jumlah kemunculan kata/term t di dalam dokumen (d)}$$

Konsep term frequency menggunakan suatu dokumen sebagai bag of words (kantong kata) di mana urutan dari kemunculan suatu kata diabaikan dan hanya jumlah kemunculan dari kata itu saja yang penting. Adapun kelemahan konsep term frequency yaitu semua kata dianggap sama. Hal ini mengakibatkan relevansi suatu kata menjadi sangat tinggi jika kata itu sering muncul dalam suatu kumpulan dokumen. Padahal tingginya frekuensi kemunculan suatu kata tidak selalu menyatakan bahwa kata tersebut penting [4].

2.4 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) menampilkan hubungan ketersediaan sebuah term dalam seluruh dokumen. Konsep Inverse Document Frequency dibuat untuk mengurangi efek dari kata frekuensinya terlalu tinggi dalam kumpulan dokumen. Ide dasarnya adalah untuk menurunkan bobot dari kata dengan frekuensi kolektif yang tinggi. Dengan kata lain, semakin banyak dokumen kata tersebut pada suatu kumpulan dokumen, maka semakin rendah bobotnya [4].

$$Idf_t = \log \frac{N}{N_t} \quad (2)$$

Keterangan:

Idf_t = Inverse Document Frequency

N = jumlah keseluruhan dokumen

N_t = jumlah dokumen yang memuat term t

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, penulis mengumpulkan data tanggapan masyarakat dari suatu media sosial yaitu twitter. Untuk memudahkan dalam penerapan algoritma untuk melakukan analisa sentimen masyarakat terhadap kinerja polri, maka penulis mengambil data sebanyak 5, data tersebut dapat dilihat pada tabel 1.

Tabel 1. Pengumpulan Data

Kode	Nama Pengguna	Tanggapan
D1	@Michellerachman4	Polri membuka lagi Kasus Dugaan Pemerkosaan Anak di Luwu Timur, Bukti Polri Tidak Diam karena Polri pengaman masyarakat
D2	@HambaAllah_411	KENAPA POLISI SULIT PROFESIONAL DAN BERKINERJA BURUK?
D3	@putri_aura_	Alhamdulillah Polri tidak berhenti, sebagai pelayanan publik polri membuka kembali kasus dugaan pemerkosaan anak di Luwu Timur karena menerima suara publik
D4	@Maryamsoloe	Polri Pengaman Masyarakat sudah sepatutnya apalagi untuk kasus pemerkosaan anak di luwu timur. Sebab sampai saat ini Polri tetap bergerak, mengkaji dan proses hukum tetap berjalan jadi tidak ada kata percuma
D5	@frankie_bjw	Polri akan membuka kembali kasus dugaan pemerkosaan tiga anak di Luwu Timur, Sulawesi Selatan. Kasus itu sempat berhenti karena kurangnya bukti. Sebagai pelayan dan pelindung masyarakat, Polri siap membuka kembali kasus tersebut jika ada bukti-bukti baru.

Setelah melakukan pengumpulan data, berikutnya tahapan yang dilakukan penulis adalah processing data:

1. Case Folding

Pada tahapan case folding yang dilakukan adalah perubahan semua huruf menjadi huruf kecil.

Tabel 2. Hasil Tahapan Case Folding

Kode	Tanggapan
D1	polri membuka lagi kasus dugaan pemerkosaan anak di luwu timur, bukti polri tidak diam karena polri pengaman masyarakat
D2	kenapa polisi sulit profesional dan berkinerja buruk?
D3	alhamdulillah polri tidak berhenti, sebagai pelayanan publik polri membuka kembali kasus dugaan pemerkosaan anak di luwu timur karena menerima suara publik
D4	polri pengaman masyarakat sudah sepatutnya apalagi untuk kasus pemerkosaan anak di luwu timur. sebab sampai saat ini polri tetap bergerak, mengkaji dan proses hukum tetap berjalan jadi tidak ada kata percuma

D5	polri akan membuka kembali kasus dugaan pemerkosaan tiga anak di luwu timur, sulawesi selatan. kasus itu sempat berhenti karena kurangnya bukti. sebagai pelayan dan pelindung masyarakat, polri siap membuka kembali kasus tersebut jika ada bukti-bukti baru.
----	---

2. Cleaning

Cleaning adalah tahapan pembersihan komponen-komponen pada dokumen yang berisi karakter atau simbol-simbol, dan angka.

Tabel 3. Hasil tahapan tokenizing

Kode	Tanggapan
D1	polri membuka lagi kasus dugaan pemerkosaan anak di luwu timur bukti polri tidak diam karena polri pengaman masyarakat
D2	kenapa polisi sulit profesional dan berkinerja buruk
D3	alhamdulillah polri tidak berhenti sebagai pelayanan publik polri membuka kembali kasus dugaan pemerkosaan anak di luwu timur karena menerima suara publik
D4	polri pengaman masyarakat sudah sepatutnya apalagi untuk kasus pemerkosaan anak di luwu timur sebab sampai saat ini polri tetap bergerak mengkaji dan proses hukum tetap berjalan jadi tidak ada kata percuma
D5	polri akan membuka kembali kasus dugaan pemerkosaan tiga anak di luwu timur sulawesi selatan kasus itu sempat berhenti karena kurangnya bukti sebagai pelayan dan pelindung masyarakat polri siap membuka kembali kasus tersebut jika ada bukti-bukti baru

3. Tokenizing

Tokenizing adalah tahap pemisahan isi teks yang semula berupa kalimat menjadi satuan kata-kata.

Tabel 4. Hasil tahapan tokenizing

Kode	Tanggapan
D1	'polri' 'membuka' 'lagi' 'kasus' 'dugaan' 'pemerkosaan' 'anak' 'di' 'luwu' 'timur' 'bukti' 'polri' 'tidak' 'diam' 'karena' 'polri' 'pengaman' 'masyarakat'
D2	'kenapa' 'polisi' 'sulit' 'profesional' 'dan' 'berkinerja' 'buruk'
D3	'alhamdulillah' 'polri' 'tidak' 'berhenti' 'sebagai' 'pelayanan' 'publik' 'polri' 'membuka' 'kembali' 'kasus' 'dugaan' 'pemerkosaan' 'anak' 'di' 'luwu' 'timur' 'karena' 'menerima' 'suara' 'publik'
D4	'polri' 'pengaman' 'masyarakat' 'sudah' 'sepatutnya' 'apalagi' 'untuk' 'kasus' 'pemerkosaan' 'anak' 'di' 'luwu' 'timur' 'sebab' 'sampai' 'saat' 'ini' 'polri' 'tetap' 'bergerak' 'mengkaji' 'dan' 'proses' 'hukum' 'tetap' 'berjalan' 'jadi' 'tidak' 'ada' 'kata' 'percuma'
D5	'polri' 'akan' 'membuka' 'kembali' 'kasus' 'dugaan' 'pemerkosaan' 'tiga' 'anak' 'di' 'luwu' 'timur' 'sulawesi' 'selatan' 'kasus' 'itu' 'sempat' 'berhenti' 'karena' 'kurangnya' 'bukti' 'sebagai' 'pelayan' 'dan' 'pelindung' 'masyarakat' 'polri' 'siap' 'membuka' 'kembali' 'kasus' 'tersebut' 'jika' 'ada' 'bukti-bukti' 'baru'

4. Stopword

Stopword adalah tahapan penghapusan kata yang tidak sesuai atau yang sering muncul. Adapun kategori stopword yaitu: ada, di, antar, aku, yang, dan lain sebagainya.

Tabel 5. Hasil tahapan stopword

Kode	Tanggapan
D1	polri buka lagi kasus dugaan pemerkosaan anak luwu timur bukti polri tidak diam karena polri pengaman masyarakat
D2	kenapa polisi sulit profesional kinerja buruk
D3	alhamdulillah polri tidak berhenti sebagai pelayanan publik polri buka kembali kasus dugaan pemerkosaan anak luwu timur karena terima suara publik
D4	polri pengaman masyarakat sudah sepatutnya apalagi untuk kasus pemerkosaan anak luwu timur sebab sampai saat polri tetap bergerak mengkaji proses hukum tetap jalan jadi tidak kata percuma
D5	polri akan buka kembali kasus dugaan pemerkosaan tiga anak luwu timur sulawesi selatan kasus itu sempat berhenti karena kurangnya bukti sebagai pelayan dan pelindung masyarakat polri siap buka kembali kasus tersebut jika ada bukti-bukti baru

5. Stemming

Stemming adalah tahapan mengubah kata-kata yang menggunakan aturan tertentu seperti menghilangkan awalan dan akhiran menjadi kata dasar.

Tabel 6. Hasil Tahapan Stemming

Kode	Tanggapan
D1	polri buka lagi kasus duga pemerkosa anak luwu timur bukti polri tidak diam karena polri pengaman masyarakat
D2	kenapa polisi sulit profesional kinerja buruk
D3	alhamdulillah polri tidak henti sebagai layan publik polri buka kembali kasus duga pemerkosa anak luwu timur karena terima suara publik
D4	polri pengaman masyarakat sudah patut apalagi untuk kasus pemerkosa anak luwu timur sebab sampai saat polri tetap gerak kaji proses hukum tetap jalan jadi tidak kata percuma
D5	polri akan buka kembali kasus dugaan pemerkosa tiga anak luwu timur sulawesi selatan kasus itu sempat henti karena kurang bukti sebagai layan lindung masyarakat polri siap buka kembali kasus tersebut jika bukti-buktinya baru

Langkah selanjutnya adalah tahap perhitungan bobot antar kata-kata tiap dokumen yang ada. Untuk memudahkan analisa ini, maka penulis menggunakan metode TF-IDF karena algoritma tersebut akan membantu penulis menganalisa antar kalimat dari setiap komentar masyarakat yang mengandung negatif, positif atau netral.

	Positif					Negatif			
	Pengaman	Profesional	Lindung	Tidak	Diam	Percuma	Buruk	Kurang	Sulit
Q	D1	D2	D3	D4	D5	df	$\frac{d}{df}$	idf	Idf+1
Pengaman	1	0	0	1	0	2	2.5	0.397	1.397
Profesional	0	1	0	0	0	1	5	0.698	1.698
Lindung	0	0	0	0	1	1	5	0.698	1.698
Tidak	1	0	1	1	0	3	1.6	0.204	1.204
Diam	1	0	0	0	0	1	5	0.698	1.698
Percuma	0	0	0	1	0	1	5	0.698	1.698
Buruk	0	1	0	0	0	1	5	0.698	1.698
Kurang	0	0	0	0	1	1	5	0.698	1.698
Sulit	0	1	0	0	0	1	5	0.698	1.698

W=tf*(idf+1)				
D1	D2	D3	D4	D5
1.397	0	0	1.397	0
0	1.698	0	0	0
0	0	0	0	1.698
1.204	0	1.204	1.204	0
1.698	0	0	0	0
0	0	0	1.698	0
0	1.698	0	0	0
0	0	0	0	1.698
0	1.698	0	0	0

Setelah menghitung bobot kata dengan algoritma TF-IDF, hasil perhitungan bobot kata yang mengandung kata positif sebesar 6.19 dan negatif sebesar 12.102. Selanjutnya menentukan persentase dari sentimen masyarakat:

$$\text{sentimen positif} = \frac{100}{6.19 + 12.102} \times 12.102 = 66.07\%$$

$$\text{sentimen negatif} = \frac{100}{6.19 + 12.102} \times 6.19 = 33.79\%$$

Dari hasil penerapan algoritma text mining dan algoritma TF-IDF pada analisis sentimen masyarakat terhadap kinerja polri, maka hasil pengujian tersebut terbukti bahwa algoritma text mining dan algoritma TF-IDF dapat menganalisa konotasi positif dan negatif pada komentar masyarakat.

4. KESIMPULAN

Berdasarkan pembobotan dari bab sebelumnya, maka dapat disimpulkan bahwa metode algoritma text mining dan TF-IDF dapat membantu dalam menghitung sentimen masyarakat. Dari hasil implementasi klasifikasi analisis sentimen terdapat presentase 66.07% sentimen positif dan 33.79% sentimen negatif.

REFERENCES

- [1] M. Hidayatullah, S. Alam, and I. Jaelani, “Sentiment Analysis of Police Performance On Twitter Users Using Naïve Bayes Method,” RISTEC Res. Inf. Syst. Technol., vol. 2, no. 2, pp. 29–30, 2021, [Online]. Available: <http://www.kontras.org/index>
- [2] N. Busiarli, L. A. Aditya, and A. Y. Andika, “Penerapan Algoritma Naïve Bayes & Natural Language Processing Untuk Mengklasifikasi Jenis Berita,” Semin. Nas. Teknol. Inf. dan Multimed., pp. 6–7, 2016.
- [3] R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge: Cambridge University Press, 2006. doi: DOI: 10.1017/CBO9780511546914.
- [4] S. Andayani and A. Ryansyah, “Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen,” JuSiTik J. Sist. dan Teknol. Inf. Komun., vol. 1, no. 1, p. 53, 2017, doi: 10.32524/jusitik.v1i1.218.
- [5] R. Feldman and J. Sanger, The Text Mining Handbook. 2006. doi: 10.1017/cbo9780511546914.