

ANALISIS PENGARUH EMOSI TERHADAP HATE SPEECH DI TWITTER: STUDI DENGAN METODE SVM DAN KORELASI

Oleh:

Muhammad Aji Prasetyo

Yulian Findawati, ST., M.MT.

TEKNIK INFORMATIKA

Universitas Muhammadiyah Sidoarjo

JULI, 2024

Pendahuluan

Kemajuan teknologi memungkinkan komunikasi cepat dan praktis melalui ponsel pintar, termasuk interaksi jarak jauh yang terjangkau. Media sosial Twitter, dengan 7,2 juta pengguna di Indonesia, menghasilkan sekitar 4,1 juta tweet harian tentang berbagai topik seperti pendidikan, hiburan, pekerjaan, dan politik. Selain untuk berkomunikasi dan mengekspresikan emosi, Twitter sering digunakan untuk menyebarkan ujaran kebencian (hate speech), terutama dalam konteks politik. Hate speech biasanya mengandung emosi negatif seperti kemarahan, yang dapat menyakiti perasaan orang lain dan memicu kekerasan. Penelitian ini bertujuan menganalisis pengaruh emosi terhadap penyebaran hate speech di Twitter menggunakan metode Support Vector Machine (SVM) untuk klasifikasi dan analisis korelasi antara emosi dan hate speech

Metode

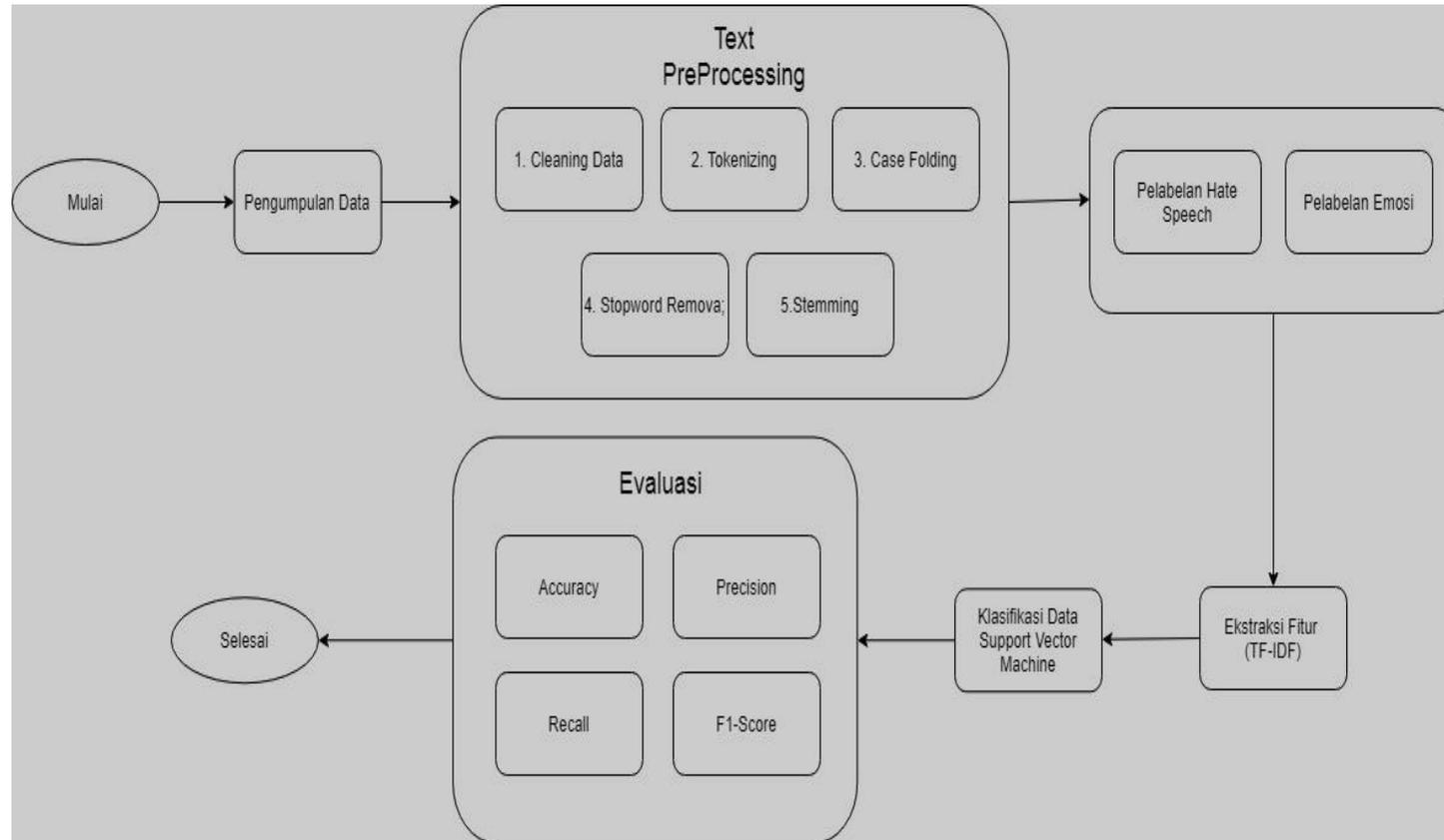


Beberapa tahapan dalam penelitian ini yang pertama melakukan pengumpulan data, data yang digunakan berasal dari akun @bumngate_tweet dengan jumlah 1500 data. Lalu dilakukan pelabelan secara manual untuk menentukan hate speech dan emosi.

Setelah dilakukan pelabelan, tahap selanjutnya adalah teks preprocessing yang tahap ini dilakukan untuk transformasi data dari yang berbentuk tidak terstruktur menjadi data yang terstruktur agar mudah untuk tahap selanjutnya.

Lalu ke tahap Pembobotan Kata (TF-IDF) dan tahap klasifikasi Support Vector Machine serta evaluasi dengan confusion matrix

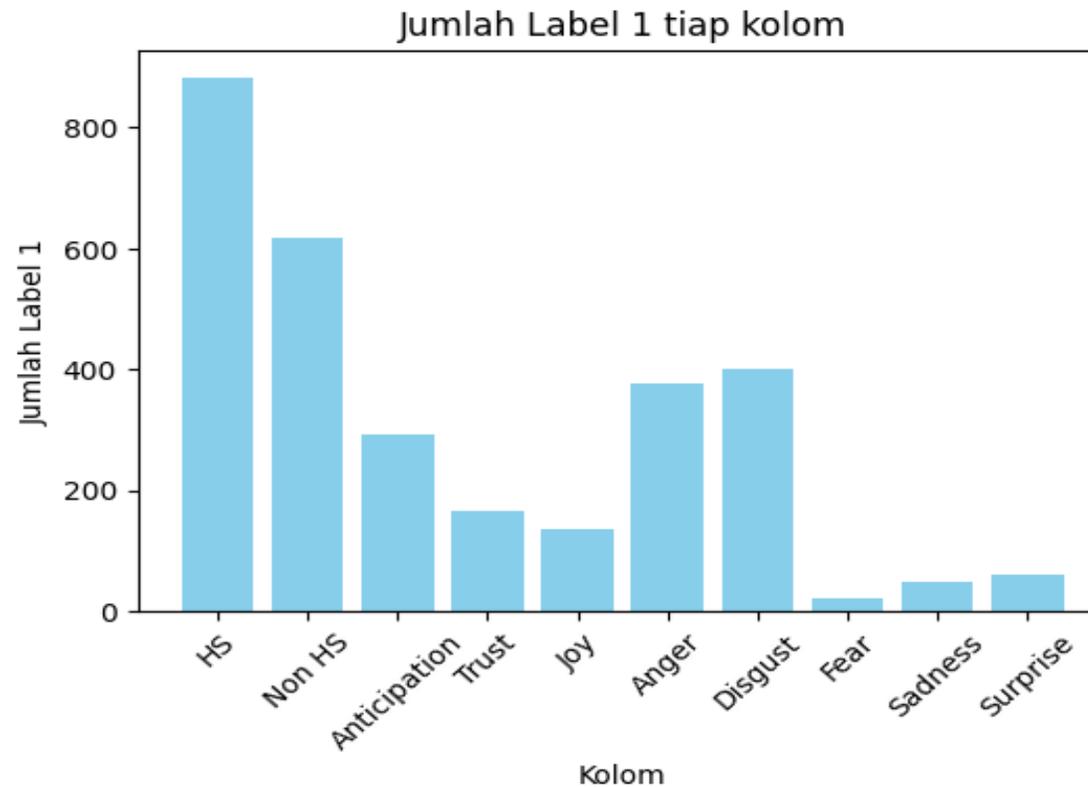
Metode



Hasil dan Pembahasan

Dataset

Dataset yang digunakan berjumlah 1500 yang telah dilabeli positif (1) dan negatif (0), berikut grafik diagram jumlah label positif (1)



Hasil dan Pembahasan

Pre-processing

Pada tahapan ini dilakukan beberapa kegiatan untuk membentuk data dan juga melakukan pembersihan data sehingga data bisa diproses ke tahap selanjutnya, Tahapan-tahapan ketika melakukan preprocessing. [11]
Sebagai berikut:

Tahapan	Hasil Text Preprocessing
Tweet	@OposisiCerdas Ha ha ha kok heran kale karena yg mau diusut gabenar yg nyapreskadrun sich, tertawa Aku termehék mehek lucu banget MERDEKA
Cleaning	OposisiCerdas Ha ha ha kok heran kale karena yg mau diusut gabenar yg nyapreskadrun sich tertawa Aku termehék mehek lucu banget MERDEKA
Case Folding	oposisicerdas ha ha ha kok heran kale karena yg mau diusut gabenar yg nyapreskadrun sich tertawa aku termehék mehek lucu banget merdeka
Tokenizing	['oposisicerdas', 'ha', 'ha', 'ha', 'kok', 'heran', 'kale', 'karena', 'yg', 'mau', 'diusut', 'gabenar', 'yg', 'nyapreskadrun', 'sich', 'tertawa', 'aku', 'termehék', 'mehek', 'lucu', 'banget', 'merdeka']
Stopword Removal	['oposisicerdas', 'heran', 'kale', 'diusut', 'gabenar', 'nyapreskadrun', 'tertawa', 'termehék', 'mehek', 'lucu', 'banget', 'merdeka']
Stemming	['oposisicerdas', 'heran', 'kale', 'usut', 'gabenar', 'nyapreskadrun', 'tawa', 'mehek', 'mehek', 'lucu', 'banget', 'merdeka']

Hasil dan Pembahasan

Pembobotan Kata

Setelah itu dataset dibagi menjadi data latih dan data uji sebanyak 80% dan 20%. Untuk pembobotan kata dengan TF-IDF, setiap kata dalam dokumen dinilai berdasarkan kemunculannya (TF) dan seberapa umum kata tersebut di seluruh korpus dokumen (IDF). Proses ini bertujuan untuk menemukan kata-kata yang paling berpengaruh dalam setiap dokumen.

Setelah dilakukan proses TF-IDF diperoleh hasil seperti pada data pertama (0, 3162) menunjukkan indeks dokumen dan indeks kata dalam kamus yang digunakan oleh "TfidfVectorizer". Sedangkan untuk 0.20533143639689172 merupakan bobot TF-IDF untuk kata tersebut dalam dokumen. Semakin tinggi bobotnya, semakin penting kata tersebut dalam dokumen tersebut

(0, 3162)	0.20533143639689172
(0, 704)	0.23669722953271066
(0, 1591)	0.5618963044053151
(0, 803)	0.5659990657268993
(0, 271)	0.21139830271886306
(0, 804)	0.4701508067186943
(1, 2154)	0.06617628938982557
(1, 3086)	0.26374631021720113
(1, 2303)	0.2685470856656998
(1, 459)	0.2685470856656998
(1, 3515)	0.27391389601489347
(1, 1376)	0.15285817366244198
(1, 2846)	0.13546158255328944
(1, 3454)	0.3370807084749853
(1, 2912)	0.13104440114996369
(1, 2733)	0.130797434564108
(1, 3688)	0.13255498888728456
(1, 1530)	0.13229972086086736
(1, 1243)	0.19323127856161212
(1, 2999)	0.14382370687744583
(1, 409)	0.1561734699051265
(1, 906)	0.41569889063228077
(1, 3095)	0.23696367943565386
(1, 778)	0.2005794977571093
(1, 2015)	0.2554387877252736
:	:
(1199, 2543)	0.44795557267879127
(1199, 1438)	0.4411090714321225

Hasil dan Pembahasan



Klasifikasi Support Vector Machine

Setelah tahap pembobotan kata dengan TF-IDF, Data akan dibagi menjadi data latih dan data uji sebesar 80% dan 20%. Selanjutnya tahap klasifikasi Support Vector Machine dengan perhitungan rumus linear kernel. Diperoleh hasil dari klasifikasi SVM dari keseluruhan semua model atau kelas dengan nilai akurasi sebesar 0.87, presisi 0.83, tetapi untuk recall hanya memperoleh 0.26 dan untuk F1-score 0.30.



Hasil dan Pembahasan

Evaluasi

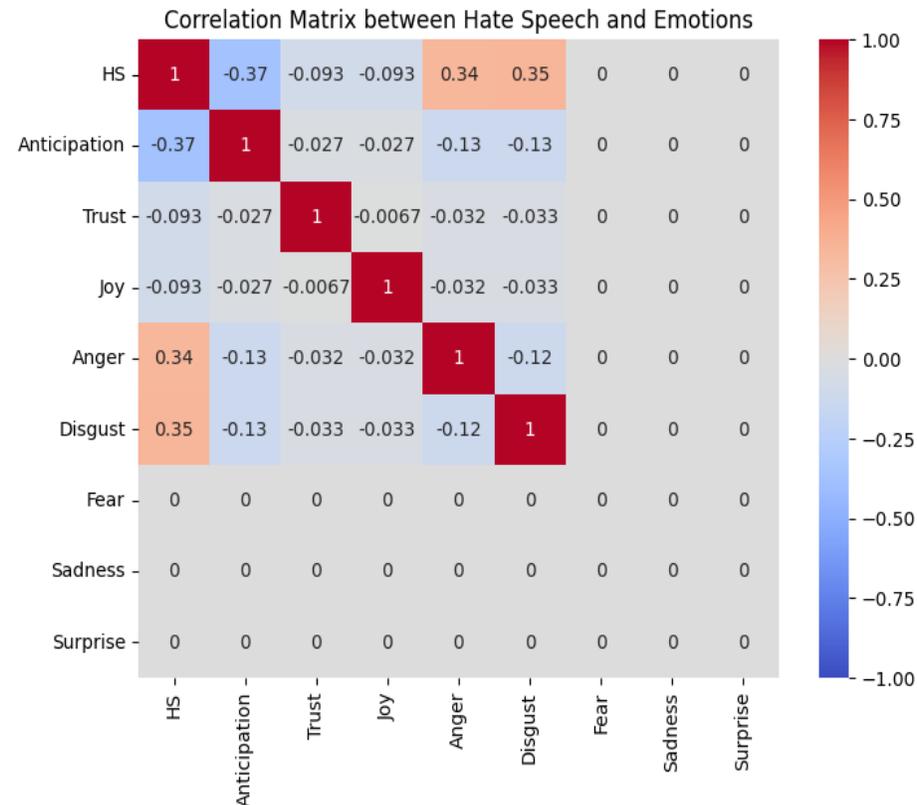
Pada tahap ini evaluasi bertujuan untuk mengukur performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan Hate Speech dan Emosi. Evaluasi ini menggunakan confusion matrix untuk menghitung akurasi, presisi, recall dan f-measure terhadap 10 kelas. Hasil confusion matrix dengan perbandingan data sebesar 80% : 20%. Berikut tabel hasil confusion matrix untuk 10 kelas.

No.	Class	TN	FP	FN	TP
1.	HS	96	22	35	147
2.	Non HS	147	35	22	96
3.	Anticipation	230	11	41	18
4.	Trust	269	0	29	2
5.	Joy	277	0	21	2
6.	Anger	207	15	54	24
7.	Disgust	202	18	57	23
8.	Fear	296	0	4	0
9.	Sadness	287	0	13	0
10.	Surprise	289	0	11	0

Hasil dan Pembahasan

Analisis Korelasi

Setelah dilakukan klasifikasi SVM pada data uji, selanjutnya kita akan menganalisis korelasi antara hate speech dan emosi. Seperti pada gambar dibawah ini.



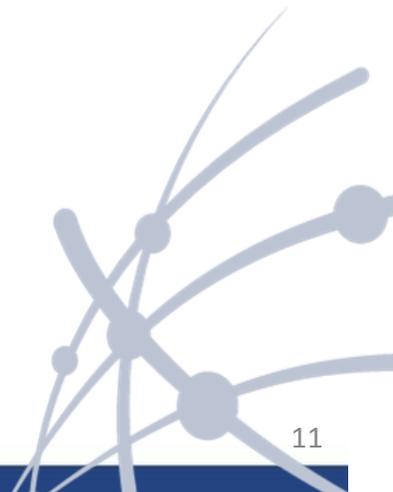
Hasil dan Pembahasan



Matriks diatas menunjukkan hubungan antara fitur hate speech (HS) dan berbagai emosi (Anticipation, Trust, Joy, Anger, Disgust, Fear, Sadness, Surprise). Dari matriks korelasi tersebut, kita dapat mengamati beberapa hal:

- a. Angka korelasi positif menunjukkan hubungan searah, sedangkan angka korelasi negatif menunjukkan hubungan berlawanan.
- b. Nilai korelasi yang mendekati 1 atau -1 menunjukkan hubungan yang kuat, sementara nilai mendekati 0 menunjukkan hubungan yang lemah atau tidak ada hubungan.

Jadi, dari matriks diatas nilai korelasi antara HS dengan Anger dan Disgust adalah tinggi, itu menunjukkan bahwa hate speech seringkali berkaitan dengan emosi marah dan jijik. Sebaliknya, jika nilai korelasi rendah atau negatif menunjukkan bahwa hate speech jarang berkaitan dengan emosi positif.



Kesimpulan

Dalam penelitian ini, data diuji dengan perbandingan 80% : 20%. Dari hasil pengujian, diperoleh rata-rata evaluasi keseluruhan kelas yaitu akurasi 87%, presisi 83%, recall 26%, dan f-measure 30%. Penelitian ini juga menghasilkan sebuah aplikasi web yang dapat mengklasifikasikan tweet ke dalam kategori hate speech atau non hate speech, serta menentukan jenis emosi yang terkandung dalam tweet, termasuk Anticipation, Trust, Joy, Anger, Disgust, Fear, Sadness, dan Surprise. Aplikasi web ini tidak hanya dapat melakukan klasifikasi, tetapi juga memungkinkan pengguna untuk menambahkan data tweet baru dan membuat laporan hasil klasifikasinya. Dari analisis korelasi yang dilakukan, ditemukan bahwa hate speech memiliki korelasi positif sedang dengan emosi Anger (Kemarahan) dan Disgust (Jijik), masing-masing sebesar 0.34 dan 0.35. Ini menunjukkan bahwa hate speech sering kali berkaitan dengan kedua emosi negatif ini. Sebaliknya, korelasi negatif sedang ditemukan antara hate speech dan emosi Anticipation (Antisipasi), dengan nilai -0.37, menunjukkan bahwa hate speech cenderung jarang berkaitan dengan emosi antisipasi.

Referensi

- [1] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [2] I. Journal, "Ijns.org Indonesian Journal on Networking and Security - Volume 11 No 2 – Juni 2022," vol. 11, no. 2, 2022.
- [3] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 334, 2019, doi: 10.22146/jnteti.v8i4.533.
- [4] J. Gaussian, "Analisis Sentimen Gojek Pada Media Sosial Twitter dengan Klasifikasi Support Vector Machine (SVM)," vol. 9, pp. 376–390, 2020.
- [5] J. K. Antartika, "Analisis Sentimen Pada Ulasan Aplikasi Home Credit dengan Metode SVM dan KNN," vol. 1, pp. 174–181, 2023.
- [6] V. M. Svm, R. W. Pratiwi, S. F. H, D. Intan, and Q. R. A, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support," vol. 8106, pp. 40–46, 2021.
- [7] R. Sistem, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan," vol. 1, no. 10, pp. 17–23, 2021.
- [8] K. I. Gunawan and J. Santoso, "Multilabel Text Classification Menggunakan SVM dan Doc2Vec Classification Pada Dokumen Berita Bahasa Indonesia," *J. Inf. Syst. Hosp. Technol.*, vol. 3, no. 01, pp. 29–38, 2021, doi: 10.37823/insight.v3i01.126.

Referensi

- [9] Ratih Puspitasari, Y. Findawati, and M. A. Rosid, "Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Classification Methods," *J. Tek. Inform.*, vol. 4, no. 4, pp. 669–679, 2023, doi: 10.52436/1.jutif.2023.4.4.801.
- [10] A. Kurniawan, Indriati, and S. Adinugroho, "Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features," *J. Pengemb. Teknol. Inf. dan Komput.*, vol. 3, no. 9, pp. 8335–8342, 2019, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6153>
- [11] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, "Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.
- [12] T. Chandra Harita, R. Kridalukmana, and D. Eridani, "Pengembangan Aplikasi Analisis Sentimen Terhadap Brand Berbasis Web Menggunakan Kerangka Kerja Flask Web-Based Sentiment Analysis Application Development Using Flask Framework," *J. Tek. Komput.*, vol. 1, no. 2, pp. 36–40, 2022, doi: 10.14710/jtk.v1i2.36307.

