

M. Aji. Prasetyo 201080200051

Artikel HKI-1.docx

by 2 Perpustakaan UMSIDA

Submission date: 05-Jul-2024 09:12AM (UTC+0700)

Submission ID: 2412667034

File name: M. Aji. Prasetyo 201080200051 Artikel HKI-1.docx (1.56M)

Word count: 3058

Character count: 19048

Analysis of the Influence of Emotions on Hate Speech on Twitter: A Study Using SVM and Correlation Methods

[Analisis Pengaruh Emosi terhadap Hate Speech di Twitter: Studi dengan Metode SVM dan Korelasi]

Muhammad Aji Prasetyo ¹⁾, Yulian Findawati ²⁾, Suprianto ³⁾, Ika Ratna Indra Astutik ⁴⁾

Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: yulianfindawati@umsida.ac.id

Abstract. In the digital era, social media platforms such as Twitter have become an integral part of Indonesian society, but they also serve as a platform for the spread of hate speech. This study aims to analyze the influence of emotions (Anticipation, Trust, Joy, Anger, Disgust, Fear, Sadness, and Surprise) on the spread of hate speech on Twitter using the Support Vector Machine (SVM) method and correlation analysis. The dataset was divided into 80% for training and 20% for testing, with stages including manual labeling, text preprocessing, word weighting with TF-IDF, and classification using SVM. The results on the test data showed an accuracy of 87%, precision of 83%, recall of 26%, and an f1-score of 30%. The resulting web application can classify tweets into hate speech or non-hate speech categories and determine the types of emotions contained in the tweets. Correlation analysis showed that hate speech has a moderate positive correlation with Anger (0.34) and Disgust (0.35), and a negative correlation with Anticipation (-0.37). This study provides important insights into the emotional dynamics in the spread of hate speech on Twitter, laying a strong foundation for hate speech classification and emotion analysis, and can be further developed to monitor and address the spread of hate speech on social media.

Keywords – Classification, Correlation Analysis, Hate Speech, Emotions, Support Vector Machine (SVM), Twitter

Abstrak. Dalam era digital, media sosial seperti Twitter telah menjadi bagian penting dari kehidupan masyarakat Indonesia, tetapi juga menjadi platform untuk penyebaran ujaran kebencian (Hate Speech). Penelitian ini bertujuan menganalisis pengaruh emosi (Antisipasi, Percaya, Senang, Marah, Jijik, Takut, Sedih, dan Terkejut) terhadap penyebaran ujaran kebencian di Twitter menggunakan metode Support Vector Machine (SVM) dan analisis korelasi. Dataset dibagi menjadi 80% sebagai data latih dan 20% sebagai data uji, dengan tahapan meliputi pelabelan manual, preprocessing teks, pembobotan kata dengan TF-IDF, dan pengklasifikasian menggunakan SVM. Hasil penelitian terhadap data uji menunjukkan akurasi sebesar 87%, precision 83%, recall 26%, dan f1-score 30%. Aplikasi web yang dihasilkan mampu mengklasifikasikan tweet ke dalam kategori hate speech atau non hate speech, serta menentukan jenis emosi yang terkandung dalam tweet. Analisis korelasi menunjukkan hate speech memiliki korelasi positif sedang dengan emosi Marah (0.34) dan Jijik (0.35), serta korelasi negatif dengan Antisipasi (-0.37). Penelitian ini memberikan wawasan penting tentang dinamika emosional dalam penyebaran hate speech di Twitter, membangun dasar yang kuat untuk klasifikasi hate speech dan analisis emosi, serta dapat dikembangkan lebih lanjut untuk memantau dan mengatasi penyebaran hate speech di media sosial.

Kata Kunci – Klasifikasi, Analisis Korelasi, Ujaran Kebencian, Emosi, Support Vector Machine (SVM), Twitter

I. PENDAHULUAN

Dalam era kemajuan teknologi yang pesat ini, komunikasi tidak hanya bisa dilakukan secara tatap muka atau secara lisan. Namun sekarang cukup dengan menggunakan ponsel pintar, maka kita bisa lebih mudah melakukan komunikasi dengan cepat dan praktis. Bahkan dengan jarak yang jauh sekalipun seperti antar pulau dan antar negara, dan juga dengan biaya yang terjangkau, komunikasi masih bisa dilakukan. Di ponsel pintar sendiri banyak aplikasi-aplikasi yang melayani komunikasi seperti Facebook, WhatsApp, Instagram, Twitter dan lain-lain.

Social media Twitter memiliki 7,2 juta pengguna di Indonesia. Setiap harinya, setidaknya terdapat 4,1 juta tweet yang dihasilkan oleh pengguna dari Indonesia (CNN Indonesia, 2016). Jumlah yang cukup besar tersebut merupakan cuitan para penggunanya tentang banyak hal: pendidikan, hiburan, pekerjaan, dan termasuk juga politik. [1] Selain itu juga pengguna memanfaatkan twitter sebagai media untuk berkomunikasi, menyatakan pendapat, dan mengungkapkan ekspresi emosi. Namun, dibalik banyak kemudahan tersebut, banyak orang yang menyalahgunakan twitter sebagai tempat untuk menyebarkan ujaran kebencian atau lebih dikenal sebagai Hate Speech.

Hate speech sendiri lebih banyak ditemukan khususnya di topik politik. Banyak orang yang melakukan ujaran kebencian kepada lawan politik hanya untuk kepentingan tertentu. Pada umumnya Hate Speech sendiri memiliki karakteristik emosi yang negatif seperti marah-marah. Hal ini jelas akan menyakitkan perasaan orang lain dan akan menimbulkan permasalahan. Dan pada akhirnya akan berujung kepada tindak kekerasan dan kejahatan.

Copyright © Universitas Muhammadiyah Sidoarjo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

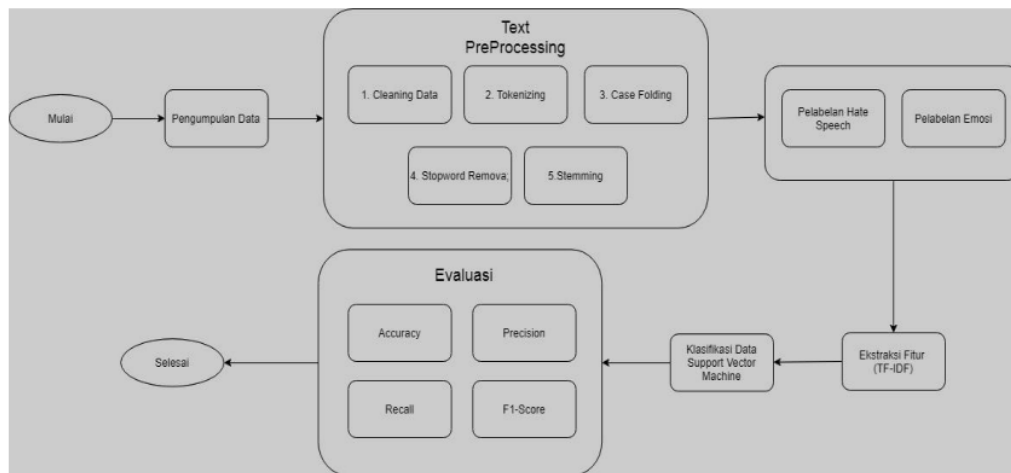
The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply

Penelitian ini bertujuan untuk menganalisis pengaruh emosi terhadap penyebaran hate speech di platform Twitter. Untuk mencapai tujuan tersebut, peneliti akan melakukan text Preprocessing, Pembobotan kata, dan menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan hate speech dan emosi. Selain itu, penelitian ini juga akan melakukan analisis korelasi untuk memahami hubungan antara emosi dan hate speech.

Support Vector Machine (SVM) didefinisikan sebagai suatu rangkaian metode pembelajaran terkait yang mengevaluasi data dan mengidentifikasi pola, yang kemudian dimanfaatkan untuk keperluan klasifikasi dan analisis regresi. SVM mengambil satu set data input dan menghasilkan prediksi untuk tiap input yang disediakan, yang berasal dari dua kelompok yang kemudian di klasifikasikan dengan mencari nilai hyperplane terbaik.[2]

II. METODE

Dalam penelitian ini, metode yang digunakan meliputi survey, layanan eksperimental untuk pengumpulan data dari Twitter. Beberapa tahapan dapat dijelaskan melalui flowchart sebagai berikut.



Gambar 1. Diagram Alir Penelitian

2.1 Pengumpulan Data

Teknik pengumpulan data menggunakan data sekunder yang diambil dari Twitter, data tweet ini berasal dari akun @bumngate_tweet, dengan jumlah 1500 data yang diambil pada tahun 2022. Kemudian dilakukan pelabelan secara manual untuk menentukan hate speech dan emosi.

Tabel 1. Contoh Dataset

No.	Tweets	HS	NHS	Ant	Trs	Joy	Ang	Dis	Fer	Sad	Sur
1.	Hahaha sigundul penguasa ancol karena selama ini taunya hanya jilat2...	1	0	0	0	0	0	1	0	0	0

Keterangan :

HS	: Hate Speech
NHS	: Non Hate Speech
Ant	: Anticipation (Antisipasi)
Trs	: Trust (Kepercayaan)
Joy	: Joy (Kegembiraan)
Ang	: Anger (Kemarahan)
Dis	: Disgust (Jijik)
Fer	: Fear (Ketakutan)
Sad	: Sadness (Kesedihan)
Sur	: Surprise (Kejutan)

15

2.2 Labelling Data

Dalam konteks pelabelan data, diputuskan untuk mengelompokkan data menjadi dua kategori utama, yaitu Hate Speech dan Non Hate Speech. Kategori Hate Speech diperuntukkan bagi ekspresi emosi seperti Anger (kemarahan), Disgust (rasa jijik), Fear (ketakutan), Sadness (kesedihan), dan Surprise (keheranan), yang secara umum terkait dengan ekspresi emosi yang mengakibatkan penderitaan. Sebaliknya, kategori Non Hate Speech digunakan untuk mencakup ekspresi emosi seperti Anticipation, Trust (kepercayaan), dan Joy (kegembiraan), yang secara garis besar berkaitan dengan nilai-nilai positif atau kebajikan. Proses pelabelan dilakukan dengan metode manual yang bersifat subjektif. [3]

3 Text Preprocessing

Proses ini merupakan tahap yang dilakukan untuk transformasi data dari yang berbentuk tidak terstruktur menjadi data yang terstruktur agar mudah untuk tahap selanjutnya. Tahapan yang dilakukan diantaranya adalah sebagai berikut:[4]

1. Cleaning merupakan penghapusan karakter yang tidak dibutuhkan seperti angka, tanda baca, emoticon, dan lain sebagainya
2. Case folding merupakan tahapan mengubah teks menjadi huruf kecil.
3. Tokenizing merupakan tahapan untuk memecah teks menjadi token atau kata-kata individual.
4. Stopword removal merupakan tahapan menghapus kata-kata umum yang sering muncul dan memiliki sedikit nilai informatif, seperti “di”, “dan”, “yang”, “dari”, “pada”
5. Stemming merupakan proses mengonversi kata-kata yang mengandung imbuhan ke dalam bentuk kata dasar[5]

2.4 Pembobotan Kata (TF-IDF)

TF-IDF merupakan pengalihan kedua algoritma yakni Term Frequency dan Inverse Document Frequency. Term Frequency adalah banyaknya jumlah term yang muncul pada suatu dokumen, sedangkan document frequency adalah pengurangan term yang sering muncul diberbagai dokumen, dengan perhitungan dari kebalikan frekuensi dokumen yang mengandung suatu kata. [6] Berikut rumus untuk IDF dan TFIDF sebagai berikut: [7]

$$IDF = \log \left(\frac{D}{DF} \right) \dots\dots\dots (1)$$

$$W = TF \times IDF \dots\dots\dots (2)$$

Keterangan:

D	: Jumlah dokumen
DF	: Frekuensi dokumen
W	: Bobot
TF	: Frekuensi kata
IDF	: Inverse document frequency

2.5 Klasifikasi Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah suatu algoritma machine learning yang mampu menangani tugas klasifikasi dan regresi. Berbeda dengan beberapa metode lainnya, SVM berfokus pada pencarian hyperplane terbaik yang mampu memisahkan dua kelas data. [8] Hyperplane ini dipilih dengan memperhatikan jarak minimum antara

hyperplane dan instance data dari masing-masing kelas adalah maksimum. SVM juga mampu menangani masalah data yang bersifat non-linear, untuk menanganinya dapat menggunakan fungsi kernel, rumusnya sebagai berikut:

Linear Kernel

Fungsi linear kernel merupakan suatu metode yang efektif ketika data telah terpisah secara linear.[9] Implementasinya dapat dilakukan melalui persamaan sebagai berikut :

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^T \cdot \vec{x}_j \dots\dots\dots (3)$$

Keterangan:

- $K(\vec{x}_i, \vec{x}_j)$: Nilai linear kernel antara x_i dan x_j
- \vec{x}_i^T : Transpose dari vector fitur x_i
- \vec{x}_j : Vektor fitur kedua

2.6 Evaluasi

Pada penelitian ini, evaluasi akan menggunakan metode matriks kebingungan (confusion matrix). Matriks kebingungan adalah tabel yang dipergunakan untuk mengevaluasi performa model klasifikasi pada suatu dataset. Berikut menunjukkan tabel confusion matrix.[10]

Tabel 2. Confusion Matrix

Kategori	Klasifikasi sebagai kategori positif	Klasifikasi sebagai kategori negatif
Positif	True Positif (TP)	False Negatif (FN)
Negatif	False Positif (FP)	True Negatif (TN)

Berdasarkan tabel yang disajikan diatas, dapat dilakukan perhitungan terhadap tingkat keakuratan (accuracy), ketepatan (precision), peringatan (recall), dan nilai f-measure.

- a Akurasi merupakan proporsi data uji yang diklasifikasikan secara tepat oleh sistem sesuai dengan kelas asalnya, dengan menggunakan persamaan sebagai berikut.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \dots\dots\dots (5)$$

- b Precision adalah nilai konsistensi antara kelas data yang dihasilkan oleh sistem dengan kelas yang sebenarnya, dengan menggunakan persamaan sebagai berikut.

$$Precision = \frac{tp}{tp+fp} \dots\dots\dots (6)$$

- c Recall adalah nilai yang menggambarkan sejauh mana sistem menghasilkan data dengan tepat berdasarkan kelas sebenarnya, dengan menggunakan persamaan sebagai berikut.

$$Recall = \frac{tp}{tp+fn} \dots\dots\dots (7)$$

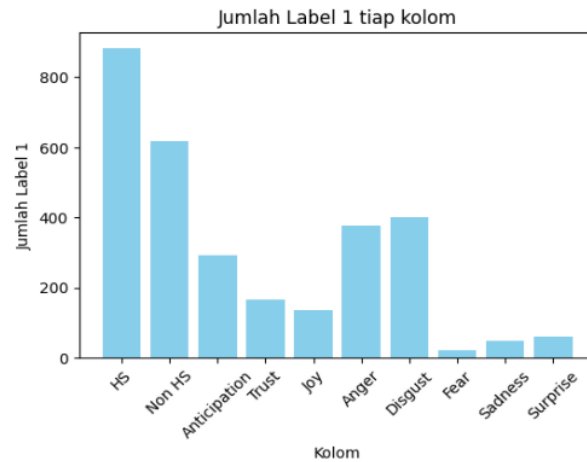
- d F-measure adalah gabungan dari precision dan recall, dengan menggunakan persamaan sebagai berikut.

$$F = \frac{2 \times precision \times recall}{precision+recall} \dots\dots\dots (8)$$

10 III. HASIL DAN PEMBAHASAN

3.1 Dataset

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



Gambar 2. Grafik Label Positif Tiap Kolom

3.2 Text Preprocessing

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:

Tabel 3. Proses Text Preprocessing

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

3.3 PEMBOBOTAN KATA (TF-IDF)

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

Gambar 3. Hasil Algoritma TF-IDF

3.4 Klasifikasi Support Vector Machine (SVM)

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.

3.5 Evaluasi

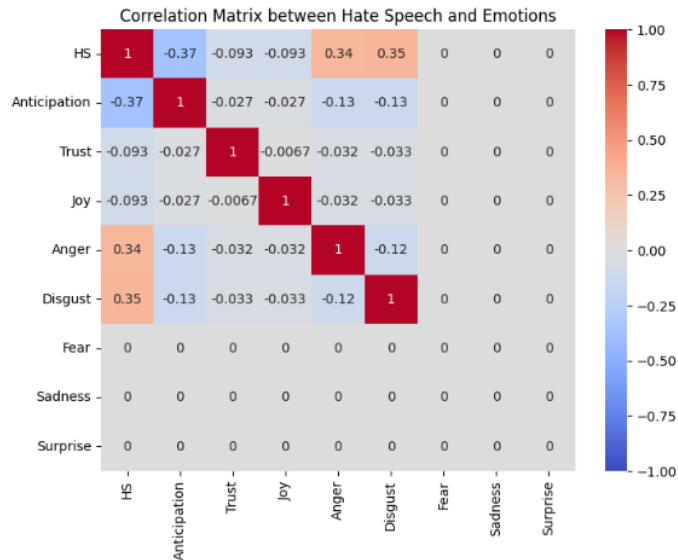
Pada tahap ini evaluasi bertujuan untuk mengukur performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan Hate Speech dan Emosi. Evaluasi ini menggunakan confusion matrix untuk menghitung akurasi, presisi, recall dan f-measure terhadap 10 kelas. Hasil confusion matrix dengan perbandingan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

Tabel 4. Confusion Matriks Tiap Kelas

No.	Class	TN	FP	FN	TP
1.	HS	96	22	35	147
2.	Non HS	147	35	22	96
3.	Anticipati on	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

3.6 Analisis Korelasi

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



Gambar 4. Matriks Korelasi HS dan Emosi

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:

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

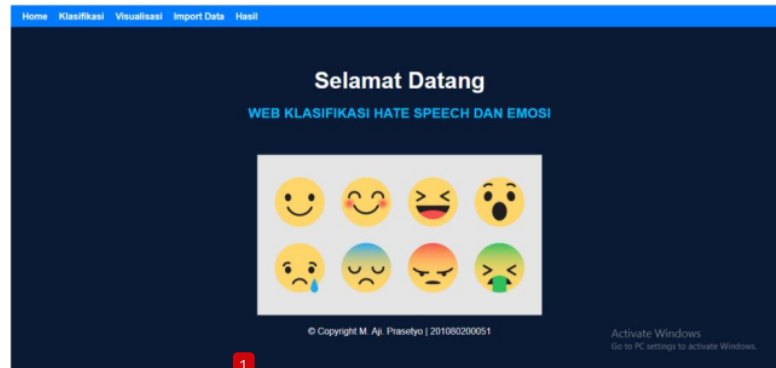
Jadi, dari matriks diatas nilai korelasi antara HS dengan Anger dan Disgust adalah tinggi, itu menunjukkan bahwa hate speech seringkali berkaitan dengan emosi marah dan jijik. Sebaliknya, jika nilai korelasi rendah atau negatif menunjukkan bahwa hate speech jarang berkaitan dengan emosi positif.

3.7 Tampilan Aplikasi

Berikut penjelasan tiap halaman pada aplikasi Klasifikasi Hate Speech dan Emosi pada Tweet berbasis Web :

1. A. Halaman Index

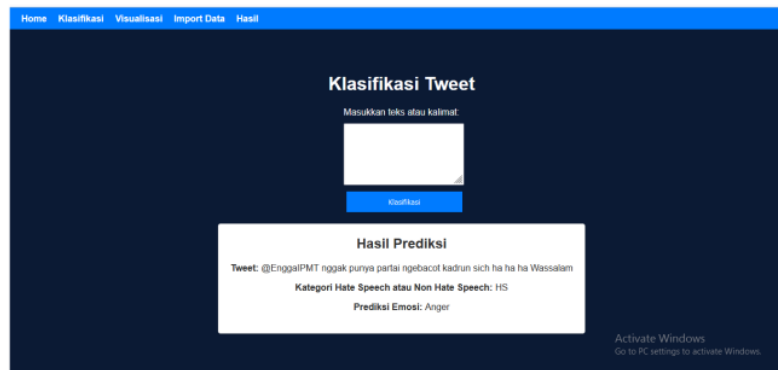
Halaman ini berfungsi sebagai tampilan awal pada aplikasi Klasifikasi Hate Speech dan Emosi berbasis Web. Berikut merupakan tampilan halaman index dibawah ini:



Gambar 5. Tampilan Halaman Index

B. Halaman Klasifikasi

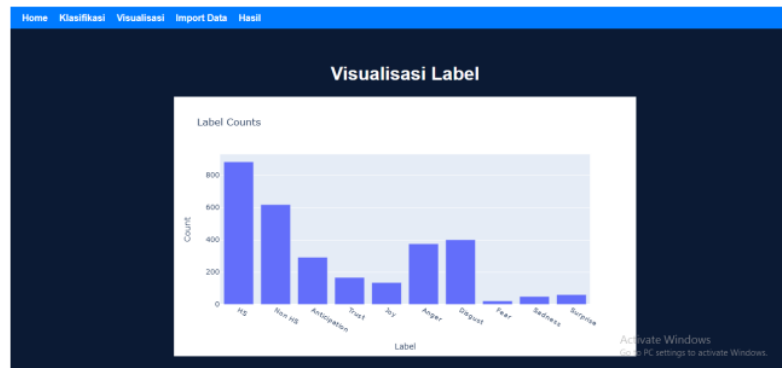
Halaman ini berfungsi untuk menguji teks atau kalimat pada tweet yang kemudian mengklasifikasikan hate speech atau non hate speech serta emosi



Gambar 6. Tampilan Halaman Klasifikasi

C. Halaman Visualisasi

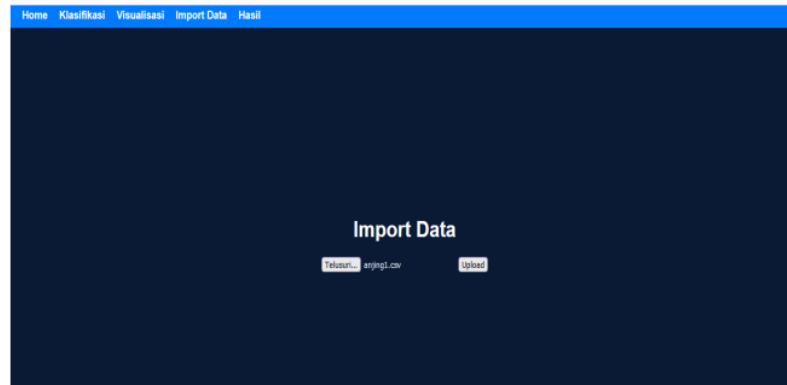
Halaman ini berisi tampilan visualisasi label positif (1) tiap kolom pada file dataset yang dipakai.



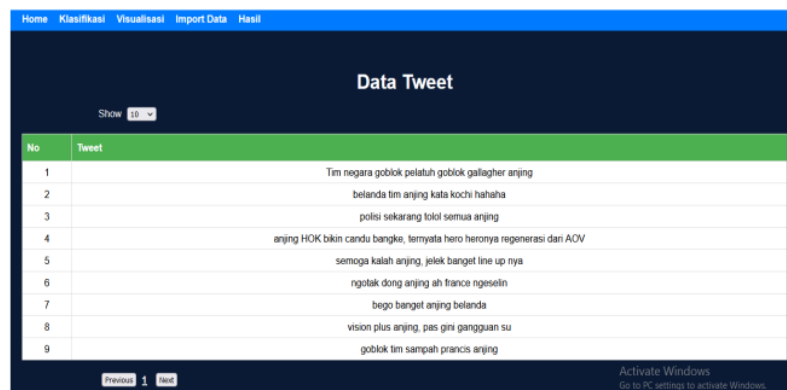
Gambar 7. Tampilan Halaman Visualisasi

D. Halaman Import Data

Halaman ini berisi import dataset yang kemudian diupload, dan menampilkan isi dari dataset tersebut.



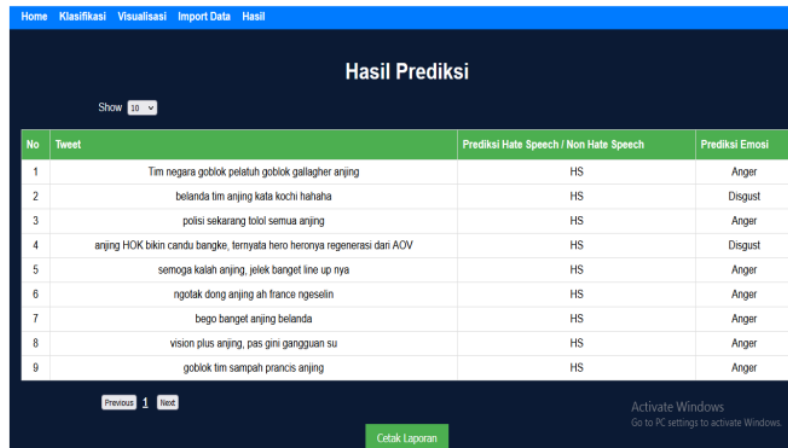
Gambar 8. Tampilan Halaman Import Data



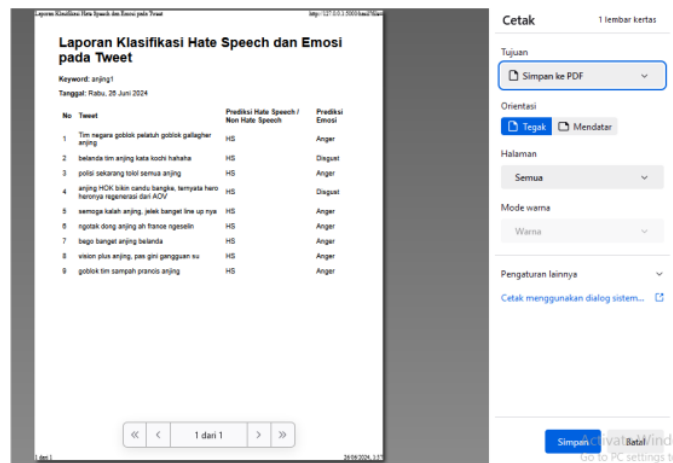
Gambar 9. Tampilan Halaman Show Data

E. Halaman Hasil

Halaman ini berfungsi menampilkan hasil prediksi klasifikasi hate speech dan emosi pada dataset yang sebelumnya sudah diupload. Lalu terdapat fitur cetak laporan berupa pdf.



Gambar 10. Tampilan Halaman Hasil



Gambar 11. Tampilan Halaman Cetak Laporan

F. Pengujian Implementasi

Metode pengujian yang dilakukan pada aplikasi klasifikasi hate speech dan emosi pada tweet adalah metode Black Box Testing. Pemilihan metode ini difokuskan untuk setiap inputan tweet dan output berupa prediksi hate speech dan emosi serta mengevaluasi hasil prediksi apakah sudah tepat atau belum. Hasil pengujian black box pada aplikasi klasifikasi hate speech dan emosi pada tweet dapat dilihat pada tabel berikut. [12]

¹ **Tabel 4. Hasil Pengujian Black Box**

No.	Kasus Uji	Keterangan
1	Memasukkan contoh kalimat tweet dengan prediksi hate speech atau non hate speech dengan tepat	Berhasil
2.	Memasukkan contoh kalimat tweet dengan prediksi emosi dengan tepat	Berhasil
3.	Dapat memasukkan file dataset dari twitter dan menampilkannya	Berhasil
4.	Dapat melakukan klasifikasi pada dataset dan menampilkannya	Berhasil
5.	Dapat mencetak laporan dari uji klasifikasi dataset	Berhasil

IV 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.

8

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada pihak-pihak yang berkontribusi dalam penelitian ini. Terimakasih kami sampaikan kepada :

1. Kedua orang tua penulis yang sudah mendukung dan memberikan semangat selama penyusunan karya tulis ini.
2. Bapak/Ibu Dosen Prodi Informatika Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo.
3. Rekan mahasiswa Universitas Muhammadiyah Sidoarjo

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.
- [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.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINALITY REPORT

15%

SIMILARITY INDEX

14%

INTERNET SOURCES

8%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	ejournal3.undip.ac.id Internet Source	2%
2	journal.ittelkom-pwt.ac.id Internet Source	2%
3	journal.untar.ac.id Internet Source	2%
4	digilib.uin-suka.ac.id Internet Source	1%
5	Submitted to Universitas Muhammadiyah Sidoarjo Student Paper	1%
6	Submitted to Troy University Student Paper	1%
7	Submitted to Universitas Brawijaya Student Paper	1%
8	pels.umsida.ac.id Internet Source	1%
9	repository.ub.ac.id Internet Source	1%

10	senafti.budiluhur.ac.id Internet Source	1 %
11	Dwi Rahma Firmansyah, Endang Lestariningsih. "Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Playstore Menggunakan Algoritma Naive Bayes", Jurnal JTİK (Jurnal Teknologi Informasi dan Komunikasi), 2024 Publication	1 %
12	ejournal.bsi.ac.id Internet Source	1 %
13	jutif.if.unsoed.ac.id Internet Source	1 %
14	Dedy Rizaldi, Ika Ratna Indra Astutik, Mochamad Alfian Rosid. "Classification of Meme Sentiment Analysis On Kaggle.com Sites To Use Support Vector Machine Algorithm", Procedia of Engineering and Life Science, 2024 Publication	1 %
15	djournals.com Internet Source	1 %
16	jifosi.upnjatim.ac.id Internet Source	1 %

Exclude quotes On

Exclude bibliography On

Exclude matches < 1%