

ANALISIS KLASIFIKASI HATE SPEECH DAN EMOSI DALAM PLATFORM TWITTER MENGGUNAKAN METODE LONG-SHORT TERM MEMORY (LSTM)

Oleh :
Nanda Yunania
Yulian Findawati
Program Studi Informatika
Universitas Muhammadiyah Sidoarjo
Juli, 2024

Abstrak

Penelitian ini bertujuan untuk mengembangkan model klasifikasi untuk mendeteksi hate speech dan emosi dalam platform Twitter dengan metode Long-Short Term Memory. Hasil eksperimen menunjukkan bahwa model LSTM mencapai akurasi sebesar 89% dalam klasifikasi Hate speech dan 71% dalam klasifikasi emosi. Hasil ini Menunjukkan potensi metode LSTM dalam tugas analisis teks dan dapat digunakan sebagai dasar untuk pengembangan sistem deteksi otomatis di platform media sosial.

Pendahuluan

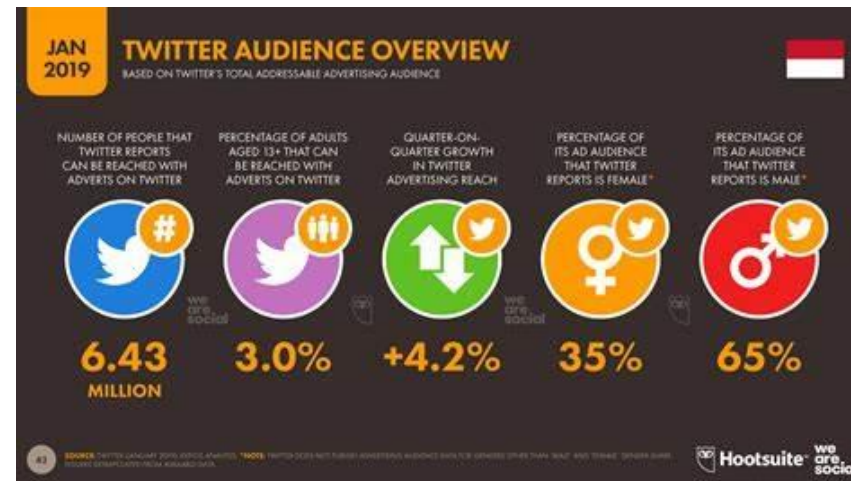
Banyaknya ragam aplikasi media sosial, banyak warga Indonesia yang menggunakan media Twitter sebagai media komunikasi dan mencari informasi.

Ujaran kebencian adalah jenis komunikasi atau ekspresi yang memiliki niat atau tujuan yang menyakiti, merendahkan kelompok atau individu berdasarkan atribut ras, etnisitas, agama, gender dan faktor lainnya

Dan juga ada ujaran emosi, merujuk pada teks yang mengandung ekspresi emosional, seperti kemarahan, kegembiraan, kesedihan atau kecemasan.

Metode

Pada penelitian ini penulis mengimplementasikan metode Long-Short Term Memory (LSTM) pada studi kasus klasifikasi hate speech dan emosi pada platform Twitter. Berikut langkah-langkah yang dilakukan :



Hasil Dan Pembahasan

► DATA SET

Himpunan data yang terdiri dari kumpulan objek dan atributnya.

| NO | Tweet |
|----------|--|
| <u>1</u> | Met Pagi manado kasih kendor' |
| 2 | Kok sewot dgn Pidato Sambutan Bpk Joko Widodo ... |
| ... | |
| 2473 | Kasihannya Ibu ini jadi korban akibat dicuci otak... |

► PREPROCESSING

Hal penting karena data pada dunia nyata terkadang tidak selalu lengkap, kotor atau noise dan tidak konsisten

| Proses | Tweet |
|-----------------------------|--|
| Case Folding | met pagi manado kasih kendor' |
| Simbol & Whitespace Removal | met pagi manado kasih kendor |
| Normalisasi Kata | selamat pagi manado jangan kasih kendor |
| Tokenizing | selamat, pagi, manado, jangan, kasih, kendor |
| Lemmatization | selamat pagi manado jangan kendor |
| Stopword Removal | Selamat manado kendor |

Hasil Dan Pembahasan

▶ SPLIT DATA

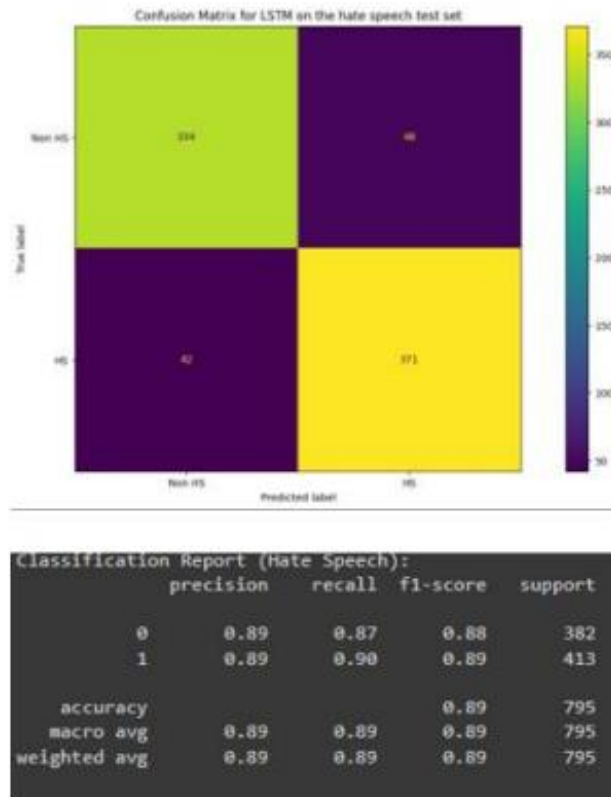
Proses membagi dataset menjadi subset-subset yang berbeda, biasanya digunakan pelatihan validasi dan pengujian model dalam pembelajaran analisis data

▶ PEMODELAN LSTM

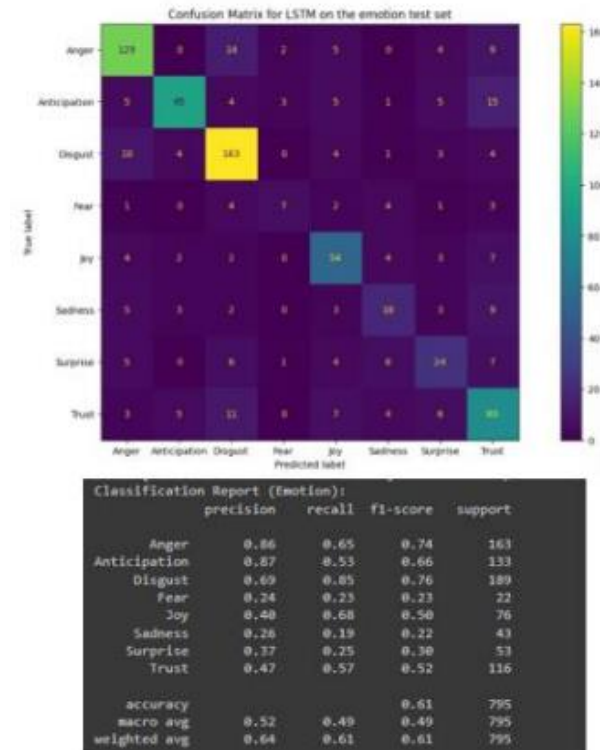
Dikembangkan untuk mengatasi masalah hilangnya informasi jangka panjang dalam jaringan saraf frekuensi konvensional. Dengan bertanya korban khusus dalam strukturnya, antarlain : forget gate, Input Gate, dan Output Gate

► KLASIFIKASI

Setelah melalui tahapan proses pembobotan kata, langkah berikutnya adalah memasukkan data ke dalam fase pemodelan klasifikasi. Sebelumnya, data akan dibagi menjadi dua bagian, yakni data latih dan data uji, dengan rasio 60:40, di mana 60% digunakan untuk data latih dan 40% untuk data uji dalam penelitian ini



Confusion Matrix Hate Speech Set



Confusion Matrix Emosi

OUTPUT

Output yang dihasilkan dari program ini adalah untuk mendeteksi Hate speech dan emosi dari sebuah kalimat random. Ada dua jenis output dari program ini yaitu berupa keterangan HS/Non HS dan Emosi

| <i>Text</i> | <i>Hate Speech</i> | <i>Emosi</i> |
|--|--------------------|--------------|
| Ha ha ha sigundul penguasa ancol | False | Trust |
| Kasihani ibu ini jadi korban akibat dicuci otak | True | Sadness |
| Kok sewot dengan pidato sambutan Bpk Joko Widodo | False | Anticipation |
| Ha ha hakadrun pada stresssssss mengenai beber | False | Trust |

KESIMPULAN

Dengan Output dari program ini adalah hasil klasifikasi teks tweet berbahasa Indonesia yang telah dianalisis menggunakan Metode LSTM. Program ini mengeluarkan dua jenis klasifikasi utama:

1. Klasifikasi Hate Speech: Program akan memberikan label pada setiap tweet apakah termasuk dalam kategori ujaran kebencian atau tidak. Ini dilakukan dengan memproses teks tweet dan membandingkannya dengan pola-pola yang telah dipelajari dari dataset pelatihan.
2. Klasifikasi Emosi: Selain mendeteksi hate speech, program juga mengidentifikasi emosi yang terkandung dalam teks tweet. Emosi yang dapat dideteksi misalnya marah, sedih, bahagia, dan lain-lain. Setiap tweet akan diberi label sesuai dengan emosi yang paling dominan berdasarkan analisis fitur linguistiknya.
3. Dapat diambil kesimpulan bahwa LSTM cukup efektif dan akurat digunakan pada klasifikasi yang menghasilkan Hate Speech 89% dan emosi 71%

Referensi

- [1] I. Liu and Y. A. Sari, "Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4914–4922, 2019.
- [2] B. A. H. Kholifatullah and A. Prihanto, "Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech," *J. Informatics Comput. Sci.*, vol. 04, pp. 292–297, 2023, doi: 10.26740/jinacs.v4n03.p292-297.
- [3] M. Murni, I. Riadi, and A. Fadlil, "Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC)," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 566, 2023, doi: 10.30865/jurikom.v10i2.5984.
- [4] E. Mardia, D. Aisha, and C. P. Dimala, "Kematangan Emosi dengan Perilaku Ujaran Kebencian Pada Remaja Akhir," vol. 11, no. 2, pp. 254–260, 2023.
- [5] J. Hartono, "Aplikasi dan Analisis Literatur Fasilkom UI," pp. 4–25, 2017, [Online]. Available: <https://123dok.com/document/yer4810q-bab-landasan-teori.html>
- [6] N. R. Radliya, "Data mining," no. 321, p. 2005, 2015.
- [7] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: ANDI, 2012.
- [8] A. A. W. Kadir, "PERBANDINGAN KINERJA KLASIFIKASI CNN BERDASARKAN STRATEGI SPLIT DATA PADA BERAGAM DATASET CITRA," 2021.
- [9] M. W. P. Aldi, Jondri, and A. Aditsania, "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin," *e-Proceeding Eng. Vol.5 No.2*, vol. 5, no. 2, pp. 3548–3555, 2018.
- [10] R. Y. Rafael and F. Adikara, "Pengimplentasian Algoritma Long Short-Term Memory Untuk Mendeteksi Ujaran Kebencian Pada Aplikasi Twitter," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 2, pp. 551–560, 2023, doi: 10.29100/jupi.v8i2.3490.
- [11] H. Henderi and R. L. Wanda, "Preprocessing Data Untuk Sistem Peramalan Tingkat Kedisiplinan Mahasiswa," *ICIT J.*, vol. 3, no. 2, pp. 296–308, 2017, doi: 10.33050/icit.v3i2.70.
- [12] N. P. S. Wati and C. Pramatha, "Penerapan Long Short Term Memory dalam Mengklasifikasi Jenis Ujaran Kebencian pada Tweet Bahasa Indonesia," *J. Nas. Teknol. Inf. dan Apl.*, vol. 1, no. 1, pp. 755–762, 2022.
- [13] I. S. Y. Saputri, M. Fadli, and I. Surya, "Implementasi E-Commerce Menggunakan Metode UCD (User Centered Design) Berbasis Web," *J. Aksara Komput. Terap.*, vol. 6, no. 2, pp. 269–278, 2017, [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jakt/article/view/1378>
- [14] Y. A. Pradana, I. Cholissodin, and ..., "Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec," ... *Teknol. Inf. dan ...*, vol. 7, no. 5, pp. 2389–2397, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12731%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/12731/5789>
- [15] C. H. Pratama and Y. Findawati, "Hate Speech and Emotions Classification in Indonesian Language Texts on Twitter Using Naïve Bayes Classifier [Klasifikasi Hate Speech dan Emosi Dalam Teks Berbahasa Indonesia Pada Pengguna Twitter Menggunakan Metode Naïve Bayes Classifier]," pp. 1–6.

