

Penggunaan K-Nearest Neighbor Untuk Klasifikasi Hate Speech Dan Emosi Pada Twitter

Oleh:

Mochamad Yanuar Kusdianto,

Uce Indahyanti

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

Februari, 2025



Pendahuluan

Perkembangan teknologi internet dan media sosial telah mengubah cara komunikasi dan interaksi manusia secara signifikan. Sebagai salah satu platform media sosial terbesar, twitter memungkinkan penggunanya untuk berbagi pendapat dan pandangan dalam bentuk tweet singkat yang dapat menjangkau audiens luas. Namun, kemudahan akses ini juga memicu penyebaran konten negatif, termasuk ujaran kebencian (hate speech). Hate speech di media sosial tidak hanya merusak tatanan sosial, tetapi juga dapat menyebabkan konflik antarindividu maupun kelompok.

Hate speech dan emosi adalah dua aspek penting dalam analisis teks di media sosial. Ujaran kebencian didefinisikan sebagai ujaran menyudutkan atau merendahkan berdasarkan atribut seperti agama, ras, atau gender. Di sisi lain, analisis emosi bertujuan untuk mengidentifikasi perasaan atau ekspresi emosional dalam teks, seperti kemarahan, kebahagiaan, atau ketakutan. Kedua jenis analisis ini semakin penting, mengingat pengaruh signifikan media sosial terhadap perilaku dan persepsi masyarakat. Penelitian yang dilakukan oleh Martins membuktikan, kemarahan dan kebencian merupakan sebuah emosi yang lebih berkonotasi dengan ujaran kebencian. Klasifikasi emosi dapat dibagi menjadi 8 yaitu terdiri dari "Fear", "Joy", "Trust", "Anticipation", "Surprise", "Sadness", "Disgust" dan "Anger".

Klasifikasi hate speech dan emosi pada Twitter menjadi semakin relevan menjelang Pilpres, karena platform ini sering digunakan sebagai arena perdebatan politik yang intens. Analisis teks menggunakan metode seperti K-Nearest Neighbor (KNN) dapat membantu mengidentifikasi ujaran kebencian (hate speech) serta emosi seperti marah, benci, atau dukungan dalam percakapan daring. Dengan meningkatnya polarisasi selama Pilpres, teknologi ini dapat digunakan untuk mendeteksi dan mengelola konten yang berpotensi memicu konflik, sehingga menciptakan ruang diskusi yang lebih sehat dan aman di media sosial.

Pada penelitian ini nantinya akan menunjukkan hasil dari klasifikasi ujaran kebencian dan deteksi emosi di sosial media twitter yang mana dalam penelitian ini metode yang dipakai ialah k nearest neighbor. Oleh karena itu, untuk mengurangi efek negatif dari interaksi di media sosial, teknologi machine learning seperti K-Nearest Neighbor (KNN) menjadi semakin penting untuk mengklasifikasikan dan mengidentifikasi hate speech serta emosi.

Pertanyaan Penelitian (Rumusan Masalah)

Berdasarkan dengan penjelasan permasalahan di latar belakang, maka dapat dirumuskan suatu masalah yakni:

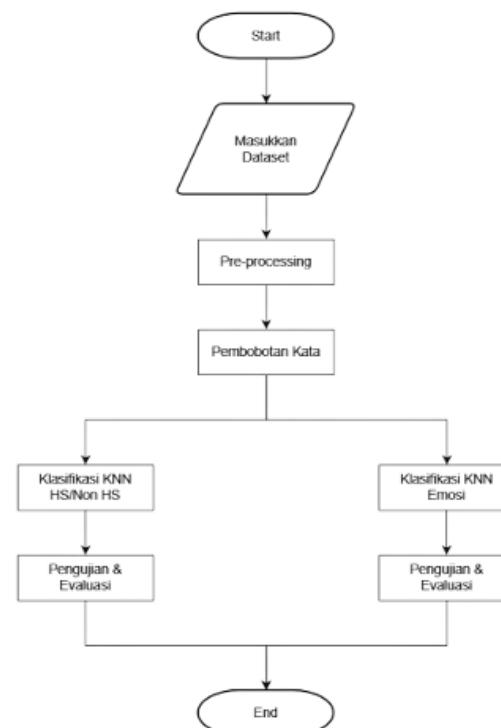
1. Bagaimana cara penggunaan K-Nearest Neighbor untuk klasifikasi hate speech dan emosi pada twitter?
2. Bagaimana keakurasiannya K-Nearest Neighbor untuk klasifikasi hate speech dan emosipada twitter?



Metode

Metodologi penelitian mencakup tahapan proses yang akan diambil untuk menyelesaikan masalah dan menemukan solusi. Penulis melakukan studi kasus dengan menggunakan K-Nearest Neighbor untuk mengkategorikan sentimen dan hate speech pada Twitter.

Penelitian ini didasarkan pada penggunaan K-Nearest Neighbor untuk mengkategorikan sentimen dan ujaran pelecehan di Twitter. Ini adalah tindakan yang dilakukan:



Hasil

Dataset yang digunakan pada penelitian ini di ambil twitter dari isu Pilpres 2024 yang berjumlah 3972 yang diklasifikasi menjadi dua yaitu dengan rasio 80% data latih dan 20% data uji. pada penelitian ini dapat dihasilkan hasil dari perbandingan antara hate speech dan non hate speech yaitu 2130 hate speech dan 1842 non hate speech. Pembobotan kata menggunakan metode TF-IDF. tujuan pada tahap pembobotan ini yaitu mengubah data berbentuk teks menjadi format numerik. Evaluasi pada penelitian ini menggunakan confision matrix yang dimana F1 Score digunakan sebagai acuan tolak ukur dalam penelitian ini. F1 score auto muncul ketika hasil klasifikasi dari presision dan recall saling bertentangan. Jika meningkatkan precision, maka kemungkinan recall akan menurun, begitupun sebaliknya F1 score digunakan untuk menemukan keseimbangan antara precision dan recall. F1 Score yaitu perbandingan rata-rata presisi dan recall dengan pemberian bobot. hasil dari evaluasi pada penelitian ini, algoritma KNN menunjukkan performa yang signifikan dalam mendeteksi hate speech dan emosi dengan akurasi rata-rata hate speech 0.8553459119496856 dan emosi 0.5345911949685535.



Pembahasan

Penelitian ini bertujuan untuk mengklasifikasikan ujaran kebencian (hate speech) dan emosi pada platform twitter menggunakan metode K-Nearest Neighbor (KNN). Adapun tahapan utama dari penelitian ini yaitu meliputi pengumpulan data melalui proses crawling, pre-processing data untuk meningkatkan kualitas data, pembobotan data menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), dan klasifikasi menggunakan algoritma KNN.

Berdasarkan hasil evaluasi, algoritma KNN menunjukkan performa yang signifikan dalam mendeteksi hate speech dan emosi dengan akurasi rata-rata 0,855 untuk klasifikasi hate speech dan 0,534 untuk klasifikasi emosi. Penelitian ini berhasil membuktikan bahwa metode KNN mampu memberikan hasil yang cukup akurat dalam menganalisis teks dengan dataset tertentu.

Temuan Penting Penelitian

Klasifikasi hate speech dan emosi menggunakan algoritma K-Nearest Neighbor, temuan pentingnya mencakup dua aspek utama. Pertama, dalam klasifikasi hate speech, model K-Nearest Neighbor berhasil mengidentifikasi tweet yang mengandung ujaran kebencian dengan akurasi yang signifikan. Hasil ini menunjukkan bahwa pendekatan K-Nearest Neighbor dapat menjadi alat yang efektif dalam memoderasi konten negatif di platform media sosial seperti twitter. Kedua, dalam klasifikasi emosi, model K-Nearest Neighbor mampu mengenali emosi yang terkandung dalam tweet dengan tingkat keakuratan yang memuaskan. Ini menunjukkan potensi algoritma ini dalam menganalisis dan memahami konten emosional dalam skala besar, yang dapat digunakan untuk berbagai tujuan, termasuk analisis sentimen dan pemahaman perilaku pengguna. Kesimpulannya, penggunaan K-Nearest Neighbor dalam klasifikasi hate speech dan emosi di Twitter menunjukkan hasil yang menjanjikan dan dapat memberikan kontribusi positif dalam upaya memahami dan mengelola konten yang beragam di platform media sosial tersebut.



Manfaat Penelitian

Penelitian klasifikasi hate speech dan emosi menggunakan algoritma K-Nearest Neighbor pada Twitter memiliki manfaat yang signifikan dalam konteks pengelolaan konten dan pemahaman perilaku pengguna. Pertama, kemampuan algoritma K-Nearest Neighbor dalam mengidentifikasi ujaran kebencian dapat membantu platform media sosial seperti Twitter untuk secara efektif memoderasi konten negatif dan mengurangi dampaknya terhadap pengguna. Ini dapat meningkatkan pengalaman pengguna secara keseluruhan dan menciptakan lingkungan online yang lebih aman dan inklusif. Kedua, kemampuan model ini dalam mengenali emosi dalam tweet dapat digunakan untuk menganalisis sentimen pengguna dan tren perilaku, yang dapat berguna bagi berbagai pihak, termasuk pemasar, peneliti, dan pengambil kebijakan, dalam membuat keputusan yang lebih tepat dan berbasis data. Dengan demikian, penelitian ini memberikan kontribusi yang berharga dalam memahami dan mengelola konten yang beragam di platform media sosial seperti Twitter.



Referensi

- [1] "PENDETEKSIAN HATE SPEECH PADA SOSIAL MEDIA INDONESIA DENGAN ALGORITMA LOGISTIC REGRESSION," pp. 1–13.
- [2] F. Poletto, V. Basile, and M. Sanguinetti, "Resources and benchmark corpora for hate speech detection : a systematic review," *Lang. Resour. Eval.*, vol. 55, no. 2, pp. 477–523, 2021, doi: 10.1007/s10579-020-09502-8.
- [3] B. Martins, G. Sheppes, J. J. Gross, and M. Mather, "Age Differences in Emotion Regulation Choice : Older Adults Use Distraction Less Than Younger Adults in High- Intensity Positive Contexts," vol. 73, no. 4, pp. 603–611, 2018, doi: 10.1093/geronb/gbw028.
- [4] I. Riadi and A. Fadilil, "Analisis Sentimen Hate Speech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC)," vol. 10, no. 2, 2023, doi: 10.30865/jurikom.v10i2.5984.
- [5] M. Alysha, Z. Larasati, N. Anisa, S. Winarsih, and M. S. Rohman, "Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter," pp. 201–210, 2016.
- [6] J. Kalyzta, M. A. Willdan, and S. Halfiani, "PENERAPAN ANALISIS SENTIMEN UJARAN KEBENCIAN TERHADAP VAKSINASI COVID-19 PADA TWEET BERBASA INDONESIA MENGGUNAKAN ALGORITME K-NEAREST NEIGHBOR," vol. 5, pp. 87–97, 2022.
- [7] R. Sistem, P. Metode, T. T. Pada, K. Teks, and H. M. K. Neighbor, "JURNAL RESTI," vol. 5, no. 10, pp. 911–918, 2021.
- [8] P. Simposium, N. Multidisiplin, and U. M. Tangerang, "Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma," vol. 4, pp. 114–121, 2022.
- [9] A. Naïve, "Jurnal KomtekInfo Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan," vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [10] R. Ariandi, U. Telkom, O. N. Pratiwi, U. Telkom, and U. Telkom, "Klasifikasi Soal Sejarah Tingkat SMA Berdasarkan Level Kognitif Revised Bloom ' s Taxonomy Menggunakan Algoritma K-Nearest Neighbour Manhattan," vol. 10, no. 2, pp. 1549–1555, 2023.
- [11] U. M. Sidoarjo, "PENGGUNAAN K-NEAREST NEIGHBOR UNTUK Oleh : Tahun 2024," 2024.
- [12] N. Aula, M. Ula, and L. Rosnita, "ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP PERUSAHAAN EKSPEDISI JNE , J & T EXPRESS DAN POS INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) ANALYSIS OF CUSTOMER REVIEW SENTIMENT TO JNE , J & T EXPRESS AND POS INDONESIA EXPEDITION COMPANIES USING SVM METHOD," vol. 9, no. 1, pp. 81–86, 2023.
- [13] L. Handayani, "IMPLEMENTASI K-NEAREST NEIGHBOR DALAM MENGKLASIFIKASIKAN HATE TWEET K-POERS PADA TWITTER Skripsi," 2023.
- [14] J. S. Komputer, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," vol. 5, no. September, pp. 697–711, 2021.
- [15] P. Romadloni, B. Adhi Kusuma, and W. Maulana Baihaqi, "Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 622–628, 2022, doi: 10.36040/jati.v6i2.5238.





DARI SINI PENCERAHAN BERSEMI