# Review Jurnal Publikasi

# “Klasifikasi Ujaran Kebencian Pada Media Sosial Twitter Menggunakan Support Vector Machine”


**Disusun Oleh :**

Rafif Ilafi Wahyu Gunawan 21081010093

**PROGRAM STUDI INFORMATIKA**
**FAKULTAS ILMU KOMPUTER**
**UNIVERSITAS PEMBANGUNAN NASIONAL**
**“VETERAN” JAWA TIMUR**


**2024**

**DETAIL PUBLIKASI**

| | | |
|---|---|---|
| **Judul Artikel** | : | Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine |
| **Nama Jurnal** | : | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) |
| **Volume** | : | Volume 5 No. 1 (2021), Halaman 17 -23 |
| **Penulis** | : | Oryza Habibie Rahman, Gunawan Abdillah, Agus Komarudin |
| **Tahun Terbit** | : | 2021 |
| **Link Artikel** | : | https://doi.org/10.29207/resti.v5i1.2700 |

## LATAR BELAKANG

Media sosial adalah bentuk dari kemajuan teknologi yang diciptakan untuk kebutuhan individu satu sama lain dalam berdampingan di kehidupan sehari - hari. Seiring perkembangan media sosial platform yang digunakan dalam kehidupan sehari - hari salah satunya twitter. Twitter adalah platform sarana bagi masyarakat untuk menyampaikan pendapat, baik itu bersifat positif maupun negatif. Terlihat dari data bahwa Indonesia merupakan salah satu negara dengan jumlah pengguna twitter terbesar, yakni mencapai 19,5 juta pengguna aktif. Salah satu permasalahan besar yang sering ditemukan di media sosial adalah ujaran kebencian (hate speech). Ujaran kebencian dapat merugikan pihak tertentu dan menjadi masalah serius, karena sekitar lima kasus ujaran kebencian yang dilaporkan setiap harinya, menandakan bahwa terdapat 150 kasus setiap bulannya.

Twitter salah satu media terbesar yang digunakan masyarakat indonesia sebagai bahan untuk mencari informasi maupun menyampaikan baik itu aspirasi maupun pendapat. Sehingga sering kali menemukan sebagian orang dalam melakukan ujaran kebencian, terlebih lagi dalam menanggapi sebuah cuitan atau konten yang sedang trending di twitter, hampir sebagian besar merupakan ujaran kebencian baik itu menyerang secara personal maupun komunal. Secara keseluruhan untuk memahami dan mengatasi ujaran kebencian di media sosial, terutama dalam konteks masyarakat Indonesia, serta bagaimana metode klasifikasi teks dapat digunakan untuk memberikan solusi dalam mengidentifikasi dan memetakan ujaran kebencian

## TUJUAN

Penelitian pada artikel ini memiliki tujuan mengidentifikasi pola perilaku masyarakat indonesia di media sosial, terutama terkait dengan ujaran kebencian. Ujaran kebencian yang sering muncul di media sosial (twitter) dapat berhubungan meliputi beberapa aspek sosial yang berbau hal - hal yang tidak jauh dari di sekitar kita misalnya suku, agama, ras, antar golongan (SARA), dan juga komentar bersifat netral. penelitian ini bertujuan untuk mengklasifikasian kategori tersebut menggunakna metode SVM yang dimana perilaku masyarakat dapat dianalisis lebih lanjut, terutama dalam hal ujaran kebencian dalam topik dan kurun waktu tertentu.

## DATASET

Data yang digunakan untuk penelitian ini dikumpulkan dari twitter dengan menggunakan sebuah “library crawling data” yang dapat mengekstraksi tweet secara otomatis. Crawling ini digunakan untuk mendapatkan dataset yang berisi tweet-tweet terkait ujaran kebencian yang dibagi ke dalam lima kelas, yaitu suku, agama, ras, antar golongan, dan netral. Proses pengumpulan ini menghasilkan 1.000 data, yang terdiri dari 700 data latih dan 300 data uji. dataset tersebut disimpan dalam format file CSV yang terdiri dari dari beberapa atribut penting seperti ID, Teks, Timestap. Berikut contoh dataset

| ID | Teks | Timestamp |
|---|---|---|
| 1 | kamu katro. Adat jawa banget. Seminggu sebelum hari H mempelai wanita di kurung gakboleh berkontak langsung dengan cowonya | 05/03/2020 21:59:33 |
| 2 | Ente khan Kristen #KAFIR, Jangan terlalu jauh lah ngurusin soal agama kami Islam, nanti ente tersesat dengan kebodohan ente sendiri. | 02/03/2020 16:58:04 |
| 3 | Pemain cina main kasar macam babi haha | 03/03/2020 12:58:29 |
| … | …… | …… |
| 999 | Siapa saja yg dukung Penista Agama adalah Bajingan yg perlu di ludahi muka nya – ADP | 01/03/2020 09:57:27 |

Gambar 1 Data Masyarakat Indonesia Bersosial Media di Twitter

## METODE PENELITIAN

Penelitian pada artikel ini menggunakan beberapa perancangan secara terstruktur dan mencakup beberapa tahap yang penting untuk memastikan keberhasilan dalam klasifikasi ujaran kebencian di twitter. berikut metode penelitian dari awal sampai akhir:

1. Pengumpulan data
    - Sumber data: Dataset yang digunakan dalam penelitian diambil dari twitter. twitter dipilih karena user pengguna aktif di indonesia termasuk banyak dan juga platform ini adalah tempat sarana seorang individu maupun kelompok menyampaikan pendapat baik secara positif maupun negatif.
    - Proses Pengumpulan Data: Dalam pengumpulan data dilakukan teknik crawling. crawling adalah proses otomatis yang digunakan untuk mengumpulkan data secara langsung dari twitter. sebuah library yang dapat melakukan crawling data digunkan untuk mengumpulkan tweet yang mengandung ujaran kebencian. dataset berhasil mencakup 1.000 tweet, yang terdiri dari 700 data latih dan 300 data uji.
2. Pengolahan data
    - Konversi Data: Data yang sudah diambil dari twitter masih belum dilabeli atau dikenal sebagai data *unsupervised*. agar dapat digunakan dalam *machine learning*, data perlu diubah menjadi data *supervised* dengan cara memberikan label pada setiap tweet
    - Klasifikasi Data: Tweet yang sudah terkumpul kemudian dilabeli secara manual oleh ditambahkan sebuah teks (annotator) yang membaca kategori dan menentukan kategori yang sesuai berdasarkan isi dari setiap tweet. Terdapat lima label atau kelas yang digunakan untuk melabeli tweet, sebagai berikut:
        1. Suku: Ujaran yang berkaitan dengan diskriminasi atau penghinaan berdasarkan suku.
        2. Agama: Ujaran yang menyudutkan atau menghina agama tertentu.
        3. Ras: Ujaran yang mengandung unsur rasisme.
        4. Antar Golongan: Ujaran yang memicu konflik antar kelompok sosial atau politik.
        5. Netral: Tweet yang tidak mengandung ujaran kebencian, melainkan hanya opini biasa.
3. Tahapan Praproses Data

Sebelum data digunakan untuk pelatihan model, data harus melalui beberapa tahap pra proses untuk membersihkan dan menstruktur data teks. Berikut merupakan langkah - langkah pra proses yang digunakan:

- Case Folding: Semua huruf dalam teks diubah menjadi huruf kecil untuk menghindari perbedaan antara huruf besar dan huruf kecil. Misalnya, "Gubernur Jawa Barat" diubah menjadi "gubernur jawa barat."
- Tokenizing: Proses ini memecah teks menjadi unit-unit kata individu. Misalnya, kalimat "silvi yudho tukang fitnah politik" dipecah menjadi ['silvi', 'yudho', 'tukang', 'fitnah', 'politik'].
- Filtering: Tahap ini bertujuan untuk menghapus kata-kata yang tidak memiliki makna penting, seperti kata penghubung atau kata-kata umum yang disebut stopwords. Misalnya, kata-kata seperti "yang", "dan", atau "dari" dihilangkan karena tidak memiliki relevansi untuk klasifikasi ujaran kebencian.
- Stemming: Proses ini digunakan untuk mengembalikan kata ke bentuk dasar atau root word. Misalnya, kata "membela" dikonversi menjadi "bela" dan "dikatakan" menjadi "kata". Stemming berguna untuk menyederhanakan variasi kata sehingga hanya kata dasar yang digunakan dalam analisis.

4. Ekstraksi Fitur dan Pembobotan
    - Setelah pra proses selesai, data teks yang sudah dibersihkan akan diproses lanjut untuk ekstraksi fitur. Sehingga metode yang digunakan yaitu term Frequency-Inverse Document Frequency (TF-IDF) untuk memberikan bobot pada setiap kata yang muncul dalam tweet. Metode ini mengukur seberapa penting sebuah kata dalam sebuah dokumen (tweet) dibandingkan dengan seluruh dataset.
    - Rumus dalam menggunakan metode TF-IDF sebagai berikut:
        - ➢ TF (Term Frequency): Mengukur seberapa sering sebuah kata muncul dalam tweet tertentu.
        - ➢ IDF (Inverse Document Frequency): Mengukur bagaimana sebuah kata didistribusikan di seluruh koleksi dokumen (tweet).
    - Setelah pembobotan menggunakan TF-IDF, data yang diperoleh dari 1.000 tweet menghasilkan 2.654 fitur (kata-kata) yang digunakan untuk klasifikasi.
5. Proses Klasifikasi dengan Support Vector Machine (SVM)

Penggunaan metode SVM berguna sebagai klasifikasi utama. SVM dipilih sebagai metode karena terbukti efektif dalam klasifikasi teks dan sering digunakan dalam penelitian serupa. Berikut mekanisme proses klasifikasi:

- Model SVM: SVM bekerja dengan mencari hyperplane yang memisahkan data ke dalam kategori-kategori yang berbeda (dalam hal ini, lima kelas: suku, agama, ras, antar golongan, dan netral). SVM pada dasarnya adalah metode klasifikasi linier, tetapi dengan penggunaan kernel yang sesuai, SVM dapat digunakan untuk memisahkan data yang non-linier.
- Teknik Kernel: Penelitian ini membandingkan tiga jenis kernel yang umum digunakan dalam SVM:
  - ➢ Linear Kernel: Kernel ini menggunakan pendekatan linier untuk memisahkan data.
  - ➢ Sigmoid Kernel: Kernel ini menggunakan fungsi aktivasi sigmoid, yang bekerja dengan baik pada data yang non-linier.
  - ➢ Radial Basis Function (RBF) Kernel: Kernel ini digunakan untuk memetakan data non-linier ke dalam ruang vektor berdimensi tinggi. Kernel RBF menunjukkan hasil terbaik dalam penelitian ini.

  Setelah fitur teks diekstraksi menggunakan TF-IDF, data dimasukkan ke dalam model SVM untuk dilatih menggunakan 700 data latih. Kemudian, model diuji menggunakan 300 data uji untuk mengukur kinerja klasifikasi.

6. Pengujian dan Evaluasi Kinerja Sistem
   - Confusion Matrix: Untuk mengukur kinerja sistem klasifikasi, digunakan metode confusion matrix yang menghasilkan beberapa metrik evaluasi kinerja, yaitu:
     - ➢ Precision: Mengukur seberapa tepat sistem mengklasifikasikan tweet ke dalam kategori yang benar.
     - ➢ Recall: Mengukur kemampuan sistem untuk menemukan semua tweet yang relevan dalam setiap kategori.
     - ➢ F-measure: Kombinasi antara precision dan recall, yang memberikan gambaran lebih baik tentang kinerja sistem secara keseluruhan.
     - ➢ Accuracy: Mengukur persentase keseluruhan tweet yang diklasifikasikan dengan benar.
   - Hasil pengujian: Dari tiga kernel yang diuji, kernel RBF menunjukkan kinerja terbaik dengan hasil akurasi 93%, precision 84%, recall 86%, dan F-measure

83%. Kernel RBF terbukti sangat baik dalam menangani data yang non-linier dan menghasilkan hyperplane yang optimal untuk klasifikasi ujaran kebencian.

7. Kesimpulan dari Pengujian Kernel

   Setelah melakukan perbandingan tiga kernel, penelitian ini menyimpulkan bahwa kernel RBF menghasilkan akurasi terbaik dalam klasifikasi ujaran kebencian. Oleh karena itu, RBF digunakan sebagai kernel utama dalam pengembangan sistem klasifikasi ini.

## HASIL & PEMBAHASAN

Hasil dan pembahasan pada artikel ini didapatkan hasil klasifikasi berdasarkan lima kelas terdiri dari suku, agama, ras, antar golongan, netral. Dimana proses klasifikasi dataset menggunkan SVM dengan beberapa kernel yang berbeda untuk memisahkan data ke dalam kategori masing - masing. Hasil klasifikasi ditujunkkan dala bentuk “confusion matrix” yang digunakan untuk mengevaluasi kinerja model dalam klasifikasi data berdasarkan beberapa metrik, yaitu (precision, recall, F-measure, dan accuracy). 3 kernel yang digunakan dalam SVM yaitu Kernel Linier, Kernel Sigmoid, Kernel RBF (Radial Basis Function). Pengujian dilakukan dengan 700 data latih dan 300 data uji menggunkan nilai parameter C = 1. Pengujian dilakukan untuk mengevaluasi kinerja setiap kernel berdasarkan “confusion matrix”, yang menghasilkan nilai precision, recall, F-measure, dan accuracy. berikut merupakan hasil dari setiap pengujian kernel

A. Kernel Linier

| Predicted Class | True Class | | | | |
|---|---|---|---|---|---|
| | Suku | Agama | Ras | Antar Golongan | Netral |
| Suku | 50 | 0 | 0 | 0 | 0 |
| Agama | 0 | 32 | 0 | 2 | 0 |
| Ras | 0 | 0 | 46 | 0 | 0 |
| Antar Golongan | 1 | 5 | 1 | 52 | 15 |
| Netral | 8 | 0 | 1 | 8 | 79 |

Tabel 1 Pengujian Kernel Linier

B. Kernel Sigmoid

| Predicted Class | True Class | | | | |
|---|---|---|---|---|---|
| | Suku | Agama | Ras | Antar Golongan | Netral |
| Suku | 47 | 0 | 0 | 0 | 0 |
| Agama | 0 | 36 | 0 | 3 | 1 |
| Ras | 0 | 0 | 45 | 1 | 0 |
| Antar Golongan | 0 | 0 | 1 | 57 | 15 |
| Netral | 4 | 0 | 1 | 6 | 83 |

Tabel 2 Pengujian Kernel Sigmoid

C. Kernel RBF (Radial Basis Funcition)

| Predicted Class | True Class | | | | |
|---|---|---|---|---|---|
| | Suku | Agama | Ras | Antar Golongan | Netral |
| Suku | 51 | 0 | 0 | 0 | 0 |
| Agama | 0 | 35 | 1 | 0 | 2 |
| Ras | 0 | 1 | 40 | 0 | 0 |
| Antar Golongan | 0 | 3 | 3 | 59 | 25 |
| Netral | 6 | 1 | 0 | 5 | 68 |

Tabel 3 Pengujian Kernel RBF

D. Tabel Hasil Pengujian Kernel

| Kernel | Average | | | F-Measure |
|---|---|---|---|---|
| | Precision | Recall | Accuracy | |
| Linear | 0.85 | 0.88 | 0.92 | 0.85 |
| Sigmoid | 0.90 | 0.89 | 0.92 | 0.87 |
| RBF | 0.84 | 0.86 | 0.93 | 0.83 |

Tabel 4 Hasil Pengujian Setiap Kernel

Dari pengujian 3 kernel didapatkan kernel RBF hasil akurasi sangat efektif dalam menangani data non-linier seperti ujaran kebencian di media sosial. RBF memiliki kemampuan untuk memetakan data ke ruang dimensi yang lebih tinggi, sehingga memungkinkan untuk pemisahan data yang lebih baik. Kinerja sistem klasifikasi diukur menggunakan beberapa metrik evaluasi sebagai berikut:

- Precision: Metrik ini mengukur seberapa tepat sistem mengidentifikasi tweet yang relevan dengan kategori tertentu. Semakin tinggi precision, semakin sedikit false positive (tweet yang salah diklasifikasikan).

- Recall: Metrik ini mengukur seberapa baik sistem menemukan semua tweet yang relevan dalam satu kategori. Semakin tinggi recall, semakin sedikit false negative (tweet yang tidak terdeteksi).
- F-measure: Kombinasi dari precision dan recall yang memberikan gambaran lebih lengkap tentang performa sistem.
- Accuracy: Mengukur persentase keseluruhan tweet yang diklasifikasikan dengan benar ke dalam kategori yang tepat.

Hasil dari evaluasi sistem menggunakan kernel RBF menunjukkan bahwa sistem klasifikasi ini memiliki performa yang cukup baik dalam mengidentifikasi ujaran kebencian di Twitter. Dengan akurasi 93%, precision 84%, recall 86%, dan F-measure 83%, sistem ini dianggap cukup efektif untuk menganalisis perilaku masyarakat dalam menyampaikan ujaran kebencian di media sosial. Oleh karena itu metode SVM dengan kernel RBF memberikan kinerja terbaik dalam klasifikasi ujaran kebencian.Hal ini dikarenakan kernel RBF memiliki kemampuan untuk menangani data yang non-linier, yang sering kali ditemukan dalam teks ujaran kebencian di Twitter. Data non-linier ini mencakup berbagai variasi bahasa, gaya penulisan, dan penggunaan bahasa informal di Twitter, yang membuat tantangan dalam klasifikasi teks menjadi lebih kompleks.

## KESIMPULAN

Kesimpulan dari artikel ini dalah bahwa Support Vector Machine (SVM), terutama dengan kernel Radial Basis Function (RBF), sangat efektif dalam mengklasifikasikan ujaran kebencian di Twitter. Penelitian ini menunjukkan bahwa sistem klasifikasi dapat membedakan tweet berdasarkan lima kategori (suku, agama, ras, antar golongan, dan netral) dengan akurasi 93%. Proses praproses data seperti case folding, tokenizing, dan stemming sangat penting untuk meningkatkan akurasi. Sistem ini dapat digunakan untuk menganalisis perilaku ujaran kebencian di media sosial dan membantu memantau serta mengelola masalah sosial yang terkait.