

Rancang Bangun Plugin Filtering Hoax Di Aplikasi Twiter Menggunakan Metode SVM

Oleh:

Agung Mulya Rasyiidi,

Cindy Taurusta

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Juli, 2024



Pendahuluan

Masyarakat sering salah artikan kebebasan berpendapat, juga dikenal sebagai kebebasan berbicara. Kebebasan berbicara, yang tercantum dalam Undang-undang Dasar Negara Republik Indonesia Tahun 1945 (UUD RI 1945). Keragu-raguan ini menyebabkan kesalahan dalam klasifikasi, yang ditakutkan akan membatasi kebebasan berbicara dan menyuarakan pendapat. Sebaliknya, membiarkan orang mengatakan apa yang mereka pikirkan tanpa batasan akan menyebabkan kebencian, diskriminasi, dan kekerasan terhadap kelompok tertentu. Media sosial adalah tempat untuk menyebarkan ujaran kebencian. Indonesia menempati peringkat ke-8 sebagai negara dengan jumlah tweet terbanyak. Tingkat kepuasan pelanggan penyedia layanan telekomunikasi seluler ditantang oleh studi analisis sentimen media sosial Twitter. Beberapa penelitian terdahulu yang menggunakan metode SVM untuk memberikan rating pada tweet berbahasa Indonesia menunjukkan akurasi 73,43%. Tujuan web menggunakan metode Support Vector Machine untuk mengklasifikasikan tweet di Twitter menjadi tweet yang menghina tokoh publik dan tweet yang tidak menghina tokoh publik. Jadi, penelitian tugas akhir ini akan membangun aplikasi analisis sentimen di Twitter, yang menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan tweet yang mengandung ujaran kebencian (hate speech) dan berita hoax.

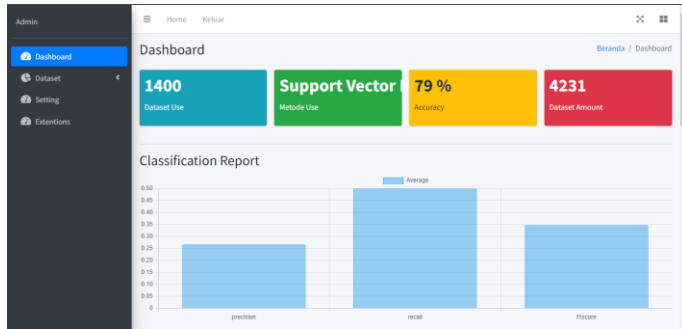
Pertanyaan Penelitian (Rumusan Masalah)

Berdasarkan latar belakang diatas maka dapat dirumuskan permasalahan dalam penelitian ini adalah bagaimana merancang dan mengembangkan aplikasi analisis sentimen di Twitter yang menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan tweet yang mengandung ujaran kebencian (hate speech) dan berita hoax?

Metode

Tahapan pertama pada penelitian ini yaitu pengambilan data dari Twitter kemudian akan diperoleh data komentar masyarakat terkait berita hoax. Setelah itu data yang diperoleh dilakukan preprocessing data. Setelah melalui pra proses data, kemudian data diberi label secara manual. Setelah itu dilakukan klasifikasi data menggunakan algoritma Support Vector Machine (SVM). Data yang sudah didapatkan akan diklasifikasikan menjadi tiga kategori kelas yaitu positif, negatif, dan netral.

Hasil



Tampilan Dashboard

The dataset page shows a table with 6 entries. Each entry includes a number, a data description, a clean status, a label, and a use train status.

No	Data	Clean	Label	Use Train	Action
1	Liter dari termometer gun akan merusak struktur otak	liter termometer gun rusak struktur otak	1	0	Action
2	Konang: Supak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	konang supak kesederhaan bisa nipu rakyat indonesia	1	0	Action
3	Gambar Denny Singar Masuh Warga Tasikmalaya di Belakang Truk	gambar denny singar masuh warga tasikmalaya belakang truk	1	0	Action
4	Foto pejabat keuarga dibawah palu arif	foto pejabat keuarga dibawah palu arif	1	0	Action
5	KPK Ditang Membawa Brimob Bersajata Masuk Gedung DPR	lpgi larang bawa brimob bersajata masuk gedung dpr	1	0	Action
6	Jembatan Soekarno Hatta di Kota Malang Melonggong	jembatan soekarno hatta kota malang melonggong	1	0	Action

Tampilan Seluruh Dataset

The dataset page shows a table with 6 entries, similar to the previous screenshot, but with a 'Use Train' column that has a value of 1 for each entry.

No	Data	Clean	Label	Use Train	Action
1	Ancaman Cina Bukan Lagi Fiksi	ancam cina bukan fiksi	0	1	Action
2	Metro TV, Malindo Perlu Ada Kode Etik Bagi Influencer	metro tv malindo perlu kode etik influencer	0	1	Action
3	FATWA SUDUT JUNAT DJALANAN	fatwa sudut junat jalan	0	1	Action
4	Ribuan Etnis China Singapura Turun ke Jalan Menentang Terpilihnya Presiden Muallam	ribu etnis china singapura turun jalan pilih presiden muallam	0	1	Action
5	Dana Haji Melayang	dana haji layang	0	1	Action
6	Worlet Dari Cina Yang Bisa Membuat Bodoh Anak	worlet cina buat bodoh anak	0	1	Action

Tampilan Penggunaan Dataset

The setting page shows a form for 'Count Data Use' with a 'Metode Use' dropdown menu. The 'Value' field contains the number 700. A 'Simpan' button is located below the form.

Tampilan Setting

The extension page shows a list of options: 'Manual' and 'Download'. The 'Extensions' menu item is highlighted in the sidebar.

Tampilan Extension

Hasil



Predicting...

<https://agung.givenjeremia.com/api/checking?keyword=Dilansir%20dari%20liputan6.com%2C%20klaim%20kabar%20Presiden%20Jokowi%20mengundurkan%20diri%20dari%20jabatan%20sebagai%20Presiden>

Tampilan Ketika Proses
Melakukan Pengujian Data



**Dilansir dari liputan6.com,
klaim kabar Presiden Jokowi
mengundurkan diri dari jabatan
sebagai Presiden**

Correct 54%

OK

Tampilan Hasil Uji Data

Pembahasan

- Pengumpulan Data

Penelitian ini menggunakan data sekunder. Dataset berisi teks berbahasa Indonesia yang diperoleh dari media sosial Twitter, dengan menggunakan bahasa pemrograman Python. Untuk melakukan steam twitter API dibutuhkan sebuah key dan access token sebagai bukti autentifikasi dengan cara developer twitter. Data yang didapat adalah seluruh opini masyarakat tentang program yang diluncurkan pemerintah. Pengumpulan data dilakukan sehingga diperoleh sebanyak 1400 data.

- Preprocessing Data

Preprocessing merupakan tahap dimana data yang diperoleh dengan cara crawling sebanyak 1400 data selanjutnya dilakukan pra-proses data. Teknik yang digunakan untuk pra-proses data yaitu Cleansing, Case Folding, Tokenizing, Filtering, Stemming.

Pembahasan

- Cleansing

Cleansing merupakan tahapan yang bertujuan menghilangkan karakter atau symbol link url (<http://situs.com>), username atau mention (@username), hastage (#), retweet, dan emoticon. Hasil data setelah melalui proses cleansing dapat dilihat pada tabel 1.

Tabel 1. Hasil Data Setelah Proses Cleansing

No	Data Mentah	Hasil Cleansing
1	RT @Kompascom: Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	kaesang bapak kesederhaan nipu rakyat indonesia
2	RT @adi098: KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	kpk larang bawa brimob senjata masuk gedung dpr
3	RT @agege: Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	acara hotman paris hutapea jadi kuasa hukum habib rizieq shihab
4	RT @slavko: Katanya Sih Mau Ngirim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	kata sih mau ngirim ketua bem ui asmat wartawan asing mau liput aja usir
5	RT @Metrotv: Kemendagri Cabut Perda Minuman Keras	kemendagri cabut perda minum keras

Pembahasan

- Case Folding

Case Folding merupakan tahapan mengganti seluruh case dalam sebuah dokumen menjadi bentuk standar (huruf kecil). Sedangkan karakter lainnya dianggap sebagai delimiter atau pembatas. Hasil data setelah melalui proses case folding dapat dilihat pada tabel 2.

Tabel 2. Hasil Data Setelah Proses Case Folding

No	Hasil Cleaning	Hasil Case Folding
1	Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	kaesang bapak kesederhaan nipu rakyat indonesia
2	KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	kpk larang bawa brimob senjata masuk gedung dpr
3	Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	acara hotman paris hutapea jadi kuasa hukum habib rizieq shihab
4	Katanya Sih Mau Ngirim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	kata sih mau ngirim ketua bem ui asmat wartawan asing mau liput aja usir
5	Kemendagri Cabut Perda Minuman Keras	kemendagri cabut perda minum keras

Pembahasan

- Tokenizing

Tokenizing merupakan tahap setelah proses case folding. Data akan di proses dimana tanda baca akan di hilangkan sehingga menghasilkan sebuah kalimat/kata yang berdiri sendiri. Hasil data setelah melalui proses tokenizing dapat dilihat pada tabel 3.

Tabel 3. Hasil Data Setelah Proses Tokenizing

No	Hasil Cleaning	Hasil Case Folding
1	Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	kaesang bapak kesederhaan nipu rakyat indonesia
2	KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	kpk larang bawa brimob senjata masuk gedung dpr
3	Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	acara hotman paris hutapea jadi kuasa hukum habib rizieq shihab
4	Katanya Sih Mau Ngrim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	kata sih mau ngirim ketua bem ui asmat wartawan asing mau liput aja usir
5	Kemendagri Cabut Perda Minuman Keras	kemendagri cabut perda minum keras

Pembahasan

- Filtering

Filtering merupakan tahapan yang bertujuan untuk menghilangkan kata umum yang biasa sering muncul dalam jumlah banyak dan tidak memiliki makna menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (Menyimpan kata penting). Hasil data setelah melalui proses Filtering dapat dilihat pada tabel 4.

Tabel 4. Hasil Data Setelah Proses Filtering

No	Hasil Tokenizing	Hasil Filtering
1	Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	kaesang bapak kesederhaan nipu rakyat indonesia
2	KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	kpk larang bawa brimob senjata masuk gedung dpr
3	Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	acara hotman paris hutapea jadi kuasa hukum habib rizieq shihab
4	Katanya Sih Mau Ngirim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	kata sih mau ngirim ketua bem ui asmat wartawan asing mau liput aja usir
5	Kemendagri Cabut Perda Minuman Keras	kemendagri cabut perda minum keras

Pembahasan

- Stemming

Stemming merupakan tahapan setiap kata akan diubah dari kata berimbuhan menjadi kata dasar. Stemming yang digunakan adalah algoritma Nazief dan Adriani yang terdapat pada library sastrawi. Hasil data setelah melalui proses Stemming dapat dilihat pada tabel 5.

Tabel 5. Hasil Data Setelah Proses Stemming

No	Hasil Filtering	Hasil Stemming
1	Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	kaesang bapak kesederhaan nipu rakyat indonesia
2	KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	kpk larang bawa brimob senjata masuk gedung dpr
3	Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	acara hotman paris hutapea jadi kuasa hukum habib rizieq shihab
4	Katanya Sih Mau Ngirim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	kata sih mau ngirim ketua bem ui asmat wartawan asing mau liput aja usir
5	Kemendagri Cabut Perda Minuman Keras	kemendagri cabut perda minum keras

Pembahasan

- Pelabelan Manual

Sebelum masuk dalam proses klasifikasi data, peneliti menggunakan pelabelan dataset manual yang melibatkan ahli bahasa dalam menentukan label tweet. Label data pada penelitian ini dibagi ke dalam tiga kategori yaitu label positif, label negatif dan label netral. Contoh hasil pelabelan manual dapat dilihat pada tabel 6.

Tabel 6. Contoh Hasil Pelabelan Manual

No	Tweet	Kelas
1	Kaesang: Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat Indonesia Kenapa Saya Tidak?	Positif
2	KPK Dilarang Membawa Brimob Bersenjata Masuk Gedung DPR	Netral
3	Pengacara Hotman Paris Hutapea Jadi Kuasa Hukum Habib Rizieq Shihab	Positif
4	Katanya Sih Mau Ngirim Ketua BEM UI ke Asmat Tapi Wartawan Asing Mau Liputan Aja di Usir	Positif
5	Kemendagri Cabut Perda Minuman Keras	Netral

Pembahasan

- Hasil Pengujian Data

Hasil yang diperoleh dari pengujian SVM dengan kernel linier dan RBF dapat dilihat pada tabel 7.

Tabel 7. Perbandingan Kernel Linier Dan RBF

Kernel	Accuracy	Precision	Recall	F1-Score
Linier	54%	0.266	0.50	0.347
RBF	79%	0.782	0.783	0.782

Kesimpulan

Penelitian tentang klasifikasi ujaran kebencian menggunakan metode support vector machine (SVM) dengan 1400 data tweet yang diambil dari media sosial twitter, dapat diambil kesimpulan bahwa:

1. Klasifikasi tweet dengan menggunakan support vector machine dapat diterapkan pada klasifikasi tweet berita hoax.
2. Parameter terbaik yang digunakan pada metode pembelajaran SVM menggunakan kernel RBF dengan pengujian confusion matrix menghasilkan akurasi 79% dengan nilai parameter $C = 1$, $\gamma = 0.01$, maksimal iterasi 1400 serta menggunakan ambang batas threshold sebesar 0.3.

Referensi

- [1] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [2] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020.
- [3] I. Aida Sapitri and M. Fikry, "Pengklasifikasian Sentimen Ulasan Aplikasi Whatsapp Pada Google Play Store Menggunakan Support Vector Machine," *J. TEKINKOM*, vol. 6, no. 1, pp. 1–7, 2023, doi: 10.37600/tekinkom.v6i1.773.
- [4] A. Muhammadin and I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc," *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 85–91, 2021, doi: 10.31294/reputasi.v2i2.785.
- [5] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.
- [6] S. D. Asri, D. Ramayanti, A. D. Putra, and Y. T. Utami, "Deteksi Roda Kendaraan Dengan Circle Hough Transform (Cht) Dan Support Vector Machine (Svm)," *J. Teknoinfo*, vol. 16, no. 2, p. 427, 2022, doi: 10.33365/jti.v16i2.1952.
- [7] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.

Referensi

- [8] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [9] E. Suryati, Styawati, and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *J. Teknol. Dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>.
- [10] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [11] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *EduTic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [12] R. A. Rizal, I. S. Girsang, and S. A. Prasetyo, "Klasifikasi Wajah Menggunakan Support Vector Machine (SVM)," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.
- [13] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [14] M. M. Maarif and N. Setiyawati, "Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
- [15] T. R. Biantong, M. T. Furqon, and A. A. Soebroto, "Implementasi Metode Support Vector Machine Untuk Implementasi Metode Support Vector Machine Untuk," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. June, pp. 185–192, 2019.

