

RESUME ARTIKEL ILMIAH

**Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu
Presiden Menggunakan Lexicon Based dan Random Forest**



Rafif Ilafi Wahyu Gunawan

21081010093

MATA KULIAH PENGOLAHAN CITRA DIGITAL A081

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL

“VETERAN” JAWA TIMUR

2024

DETAIL PUBLIKASI

Judul Artikel	:	Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest
Nama Jurnal	:	Jurnal Ilmiah Informatika
Volume	:	Vol. 11 No. 02 (2023)
Penulis	:	Oktaviامي Manullang, Cahyo Prianto, Nisa Hanum Harani
Tahun Terbit	:	2023
Link Artikel	:	10.48550/arXiv.1909.11942

LATAR BELAKANG

Pemilihan Presiden merupakan salah satu momen penting dalam demokrasi suatu negara, termasuk Indonesia. Di era digital, media sosial menjadi sarana utama bagi masyarakat untuk mengekspresikan pandangan dan opini politik mereka. Platform seperti Twitter sering digunakan untuk membahas calon presiden, dengan tweet yang mencakup sentimen positif, negatif, dan netral. Namun, banyaknya data opini yang tersebar memerlukan analisis otomatis untuk memahami pola sentimen masyarakat. Penelitian ini menggunakan analisis sentimen berbasis metode *Lexicon Based* dan *Random Forest* untuk mengevaluasi opini masyarakat terhadap calon presiden. Analisis ini bertujuan membantu kandidat dan partai politik dalam merumuskan strategi kampanye yang lebih efektif.

ISI ARTIKEL

Penelitian ini mengaplikasikan analisis sentimen untuk memprediksi preferensi publik terhadap tiga kandidat calon presiden Indonesia (Prabowo Subianto, Anies Baswedan, dan Ganjar Pranowo). Data dikumpulkan dari tweet berbahasa

Indonesia menggunakan kata kunci terkait dengan ketiga calon tersebut. Proses analisis melibatkan *text preprocessing*, pelabelan manual, dan penerapan metode *Lexicon Based* serta *Random Forest*. Penelitian ini bertujuan untuk mengevaluasi akurasi kedua metode dalam mengklasifikasikan sentimen positif, negatif, dan netral.

DATASET

Dataset terdiri dari **3.862 tweet** yang dikumpulkan melalui API Twitter menggunakan *tweet-harvest*. Data yang digunakan adalah tweet publik yang relevan dengan kata kunci seperti "#prabowo", "#anies", dan "#ganjar". Dataset ini disimpan dalam format CSV dan melalui proses *preprocessing*, termasuk:

- *Cleaning* (penghapusan URL, username, hashtag).
- *Case folding* (mengubah semua huruf menjadi huruf kecil).
- *Tokenization* (memisahkan kalimat menjadi kata-kata individu).
- *Normalization* (menyesuaikan kata dengan bentuk baku).
- *Stopword removal* (menghapus kata-kata yang tidak memiliki makna penting).

METODELOGI

Penelitian dilakukan berdasarkan lima tahapan utama sebagai berikut:

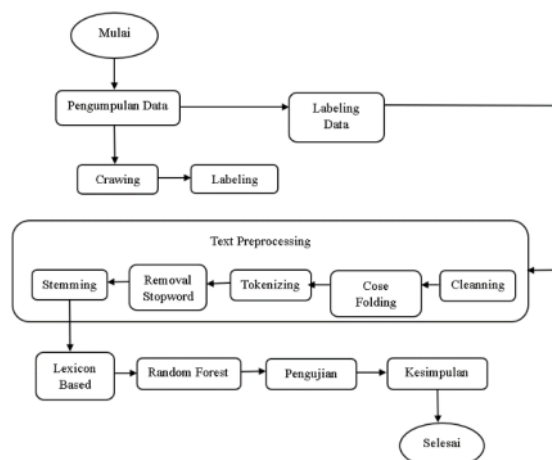


Figure 1 Alur Penelitian

1. Data Collection: Pengumpulan data menggunakan API Twitter.
2. Preprocessing: Membersihkan dan mempersiapkan data untuk analisis.
3. Text Labeling: Pelabelan manual dibantu ahli bahasa untuk menentukan sentimen positif, negatif, atau netral.
4. Sentiment Analysis:
 - *Lexicon Based*: Menggunakan kamus sentimen untuk mengklasifikasikan teks.
 - *Random Forest*: Algoritma *machine learning* berbasis pohon keputusan untuk prediksi sentimen.
5. Evaluation: Pengukuran akurasi menggunakan *Confusion Matrix*.

HASIL DAN PEMBAHASAN

Penelitian ini berhasil menganalisis sentimen publik terhadap tiga kandidat calon presiden Indonesia, yaitu Prabowo Subianto, Anies Baswedan, dan Ganjar Pranowo, berdasarkan data yang diambil dari Twitter. Dataset yang digunakan berisi 3.862 tweet dalam bahasa Indonesia, yang dikumpulkan menggunakan kata kunci seperti #prabowo, #anies, dan #ganjar. Setelah melalui tahap preprocessing dan pelabelan, tweet-tweet tersebut diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral.

Dari analisis data, ditemukan bahwa kandidat Prabowo memiliki jumlah sentimen positif tertinggi, yaitu 1.108 tweet, diikuti oleh 968 tweet negatif. Untuk Anies, ditemukan 1.754 tweet negatif dan hanya 304 tweet positif. Ganjar mendapatkan 628 tweet positif dan 1.439 tweet negatif. Hal ini menunjukkan bahwa opini publik terhadap Prabowo cenderung lebih positif dibandingkan dengan dua kandidat lainnya.

Hasil klasifikasi menunjukkan bahwa metode *Lexicon Based* memberikan akurasi sebesar 88%, dengan distribusi sentimen negatif sebesar 48%, sentimen positif 96%, dan sentimen netral 97%. Di sisi lain, metode *Random Forest* menunjukkan akurasi yang lebih tinggi, yaitu 94%, dengan hasil yang lebih stabil dan generalisasi lebih baik dibandingkan *Lexicon Based*. Hal ini menegaskan bahwa metode berbasis pembelajaran mesin seperti *Random Forest* lebih mampu

menangkap pola sentimen dalam data teks yang kompleks dibandingkan pendekatan berbasis leksikon.

Visualisasi data menggunakan wordcloud menunjukkan kata-kata yang sering muncul dalam setiap kategori sentimen. Untuk sentimen positif, kata-kata seperti "dukungan", "pemimpin", dan "program" dominan, sementara pada sentimen negatif ditemukan kata-kata seperti "gagal", "tidak setuju", dan "kritik". Analisis ini membantu memberikan wawasan tentang alasan di balik sentimen publik terhadap masing-masing kandidat.

Selain itu, analisis distribusi data menggunakan diagram batang dan pie chart menunjukkan bahwa 57,6% masyarakat secara eksplisit menyatakan dukungan terhadap salah satu kandidat, sementara sisanya, 42,4%, tidak memberikan dukungan kepada kandidat manapun. Hal ini menggambarkan bahwa, meskipun ada polarisasi opini terhadap kandidat tertentu, sebagian besar masyarakat cenderung menunjukkan sikap netral atau abstain.

Pengujian model dilakukan menggunakan Confusion Matrix, yang memberikan gambaran tentang performa klasifikasi. Pada metode Random Forest, nilai Precision, Recall, dan F1-Score untuk sentimen negatif mendekati sempurna (mendekati 1,00), menunjukkan kemampuan yang sangat baik dalam mengenali tweet dengan sentimen negatif. Namun, terdapat beberapa kesalahan prediksi, seperti sentimen netral yang salah diklasifikasikan sebagai negatif. Hal ini menunjukkan bahwa meskipun Random Forest unggul dalam akurasi, masih ada ruang untuk perbaikan, terutama dalam mengklasifikasikan sentimen yang lebih ambigu.

Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam memahami opini publik terhadap calon presiden melalui analisis sentimen. Temuan ini tidak hanya relevan bagi dunia akademik, tetapi juga dapat digunakan oleh partai politik dan tim kampanye untuk merancang strategi yang lebih responsif terhadap aspirasi masyarakat. Hasil penelitian menunjukkan bahwa pendekatan berbasis Random Forest memiliki potensi besar untuk digunakan dalam analisis data teks berskala besar di masa mendatang.

KESIMPULAN

Penelitian ini menunjukkan bahwa analisis sentimen menggunakan metode *Lexicon Based* dan *Random Forest* dapat memahami opini publik terhadap calon presiden melalui data Twitter. Kandidat Prabowo Subianto memperoleh lebih banyak sentimen positif dibandingkan Anies Baswedan dan Ganjar Pranowo, yang didominasi sentimen negatif. Metode *Random Forest* memiliki akurasi lebih tinggi (94%) dibandingkan *Lexicon Based* (88%) dan lebih andal dalam mengenali sentimen, meskipun masih terdapat kesalahan pada data ambigu. Hasil ini memberikan wawasan bagi tim kampanye dan partai politik untuk menyusun strategi berbasis data. Penelitian ini relevan untuk menganalisis opini publik pada berbagai isu, dengan potensi pengembangan lebih lanjut melalui integrasi data dari platform lain dan metode yang lebih canggih untuk meningkatkan akurasi.