

IMPLEMENTASI ALGORITMA K-MEANS TERHADAP OPINI MASYARAKAT MENGENAI PERKIRAAN PEMILU 2024 PADA TWITTER

Yepin Andrianus^{*1}, Wasino², Tri Sutrisno³
^{1,2,3}Sistem Informasi, FTI, Universitas Tarumanagara
Jl. Letjen S Parman No 1, Jakarta 11440 Indonesia

e-mail: ^{*1}yepin.825180081@stu.untar.ac.id, ²wasino@fti.untar.ac.id, ³tris@fti.untar.ac.id

Dengan adanya pemilihan umum rakyat dapat ikut berperan serta dalam memilih pemimpin. Banyak dari masyarakat yang beropini mengenai pemilihan umum melalui media sosial, terutama twitter. Dalam hal ini bisa dilakukan penelitian yang membahas mengenai opini masyarakat pada pemilihan umum. penelitian ini menggunakan algoritma k-means clustering dengan metode pengujian silhouette coefficient. Data yang diperoleh sebanyak 1000 data. Berdasarkan hasil perhitungan bahwa sentimen positif memiliki nilai 4.30%, netral 92.50% dan negatif 3.20% Dengan tingkat akurasi menggunakan metode Silhouette sebesar 0.9812638% yang memiliki hasil struktur yang kuat.

Kata Kunci: Algoritma K-Means, Data Mining, Analisis Sentimen, Pemilihan Umum

I. PENDAHULUAN

Indonesia menganut bentuk pemerintah demokrasi. Demokrasi sendiri memiliki paham dimana seluruh warga negara tanpa memandang jenis kelamin, ras, suku, agama, latar belakang maupun status sosial berhak mendapatkan hak yang sama dalam hidupnya dan dalam mengambil keputusan[1].

Salah satu bentuk nyata dari paham demokrasi ialah Pemilihan Umum. Komisi Pemilihan umum sudah menetapkan 14 Februari 2024 akan menjadi tanggal berlangsungnya pemilihan umum[2]. Walaupun pemilu belum dilaksanakan, namun banyak dari masyarakat yang sudah memberikan opini mereka melalui media sosial *Twitter*. Tujuan dilakukan penelitian ini yaitu untuk mengukur tingkat akurasi menggunakan algoritma yang berbeda dan mengetahui persentase sentimen masyarakat terhadap kasus pemilihan umum 2024.

Penelitian pertama yang dilakukan oleh Raihan dkk. Penelitian ini membahas mengenai Calon Presiden 2024 dengan menggunakan algoritma *Naive Bayes* dan menggunakan *Area Under Curve (AUC)* sebagai penentuan tingkat kualitas model. Berdasarkan hasil yang didapatkan bahwa masyarakat cenderung memiliki sentimen positif untuk masing-masing calon Presiden dengan tingkat kualitas model yang berbeda-beda[2].

Penelitian kedua yang dilakukan oleh Putu dkk. Pada penelitian ini membahas Opini Masyarakat melalui media

sosial *Twitter* dengan kata kunci "Figure Pemimpin" / "Calon Pemimpin" dengan menggunakan *Python*. Hasil yang didapatkan sebesar 21,6% sentimen positif, sentimen netral 78% dan 0,4% sentiment negatif[1].

Penelitian ketiga dan terakhir oleh Rengga dkk. Pada penelitian ini membahas Opini masyarakat terhadap Pemilihan umum 2019 melalui media sosial *Twitter* dengan menggunakan algoritma *Naive Bayes*. Data yang akan diolah berjumlah 1000 data. Berdasarkan hasil penelitian menyatakan bahwa sentimen positif memiliki persentase sebesar 22,66%, sentimen netal memiliki persentase sebesar 37,95% dan sentimen negatif memiliki persentase sebesar 39,39% serta tingkat akurasi dengan menggunakan algoritma *Naive Bayes* sebesar 62% dengan hasil *recall* sebesar 45% dan *presisi* sebesar 41%[3]

II. METODE PENELITIAN

A. Analisis Sentimen

Analisis Sentimen adalah sebuah metode yang digunakan untuk memberikan penilaian terhadap suatu kasus ataupun masalah yang dituju dari masyarakat. Analisis sentimen menggunakan metode analisis teks. Analisis sentimen memiliki 3 kategori yang terdiri dari Sentimen Positif yang melambangkan bahwa individu sepakat dengan situasi yang terjadi, Sentimen Negatif yang melambangkan bahwa individu tidak sepakat dengan situasi yang terjadi, dan Sentimen Netral melambangkan bahwa individu memiliki respon yang tidak bersifat positif maupun negatif[4].

B. Preprocessing

Preprocessing merupakan sebuah metode yang dilakukan untuk membersihkan data yang tidak rapi ataupun tidak terstruktur. Dalam proses *preprocessing* ada beberapa tahapan yang harus dilakukan yang meliputi[5]:

1. **Cleaning**: Proses menghilangkan simbol, tanda baca, link url, angka dan emoji dalam sebuah data.
2. **Case Folding**: Proses pengubahan semua huruf besar yang ada dalam data menjadi huruf kecil.
3. **Stopword Removal**: Proses penghapusan kata penghubung yang tertera dalam kamus.
4. **Tokenizing**: Proses pemecahan setiap masing-masing kata.

5. **Stemming**: Proses mengubah setiap kata menjadi kata yang baku sesuai dengan bahasa yang baik dan benar[6].

C. *Algoritma K-Means*

Algoritma K-Means adalah suatu teknik pengelompokan data dalam data mining yang dapat menghasilkan sekelompok data yang jumlahnya besar dengan cara partisi yang berdasarkan titik, dan memiliki waktu komputasi yang efisien dan cepat. Data tersebut dipilah membentuk beberapa bagian dengan parameter yang telah ditentukan lalu ditentukan kemudian disatukan menjadi satu dalam cluster. Berikut merupakan Langkah-Langkah dalam menggunakan algoritma *K-Means Clustering*

1. Menetapkan jumlah *cluster* secara sembarangan.
2. Menghitung seberapa jauh setiap masing-masing data dari pusat *cluster*.

Untuk melakukan perhitungan jarak terdekat dapat menggunakan rumus *Euclidean Distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

- d = jarak pada x dan y
- x = data pusat cluster
- y = data atribut
- i = setiap data
- n = jumlah data
- x1 = data pusat klaster ke i
- y1 = data pada setiap data ke i.

3. Menghitung pusat *cluster* baru setelah data dikelompokan melalui jarak terdekat.
4. Menghitung pusat group baru untuk setiap anggota kelas menggunakan rata-rata untuk setiap objek dari suatu kelas yang spesifik. Untuk melakukan perhitungan *centeroid* dari rata rata yang bersangkutan dapat menggunakan rumus berikut

$$c_j = \frac{1}{N_k} \sum_j^{N_k} = 1x_j$$

Keterangan:

- cj*= *Centroid* Baru
- Nk*= Jumlah keseluruhan data yang tergabung dalam *cluster*
- j*= Data kriteria *centroid* pada *cluster* ke-*j*

5. Lakukan perulangan pada tahap 2 sampai tahap 4 sampai hasil dari data tersebut tidak berpindah ke *cluster* lain[5].

D. *Silhouette Coefficient*

Setelah melakukan proses *clustering*, maka tahap selanjutnya yaitu proses pengujian. Proses pengujian ini berguna agar mengetahui tingkat akurasi dari algoritma yang digunakan. Dalam proses ini tahap pengujian menggunakan perhitungan *Silhouette Coefficient*. berikut merupakan proses perhitungan untuk menghitung nilai *silhouette coefficient*:

1. Melakukan perhitungan jarak menggunakan rata-rata nilai objek(i) dan semua titik yang ada dalam cluster.
2. Melakukan perhitungan jarak menggunakan rata-rata nilai objek(i) dan semua titik yang ada dalam cluster yang berbeda, lalu ambil nilai yang paling kecil.
3. *Silhouette coefficient* dapat ditemukan dengan menggunakan rumus berikut

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Keterangan:

a(i)= rata- rata jarak untuk setiap titik data dalam cluster pada data ke(i)

b(i) = rata-rata jarak untuk setiap cluster pada data ke(i).

Berikut merupakan tabel nilai akurasi dengan menggunakan *silhouette coefficient*[7]

III. HASIL PEMBAHASAN

A. *Data*

Untuk mendapatkan data opini masyarakat, maka perlu dilakukan *crawling* data menggunakan *tools Rapidminer*. *Rapidminer* digunakan untuk menghubungkan antara *API Twitter* dengan *Rapidminer* sendiri. Untuk mendapatkan *API twitter* sendiri pengguna harus mendaftarkan ke *website twitter developer*. Data yang telah dilakukan *crawling* akan bertipe *file csv*. Jumlah data yang diambil sebanyak 1000 data. Proses penarikan data dilakukan di bulan februari 2023 hingga bulan maret 2023. Setelah melakukan proses *crawling* data maka akan masuk ke dalam tahapan *preprocessing data*[7].

Tabel 1 Hasil *Crawling* Data

No	Tweets
1.	Mari kita sukseskan Pemilu 2024 yang AMAN, DAMAI dan SEJUK di Kota Magelang
2.	Semoga Pemilu 2024 berjalan lancar dan Jurdil. Tak ada lagi kampanye hitam, tak ada lagi politik identitas, tak ada lagi provokator. Sehingga dapat terwujudkan Pesta demokrasi yang berkualitas.
3.	Dengan penuh integritas, Kepala Kantor Imigrasi Kelas I TPI Gorontalo memimpin pembacaan ikrar diikuti oleh seluruh pegawai, berisi empat poin penting pernyataan netralitas dalam menyukseskan Pemilu 2024.
4.	Stop politik identitas! Pemimpin itu punya visi bukan memprovokasi #pemilu2024
5.	Ini dia calon pemimpi di pemilu 2024 yang suka bikin jalanan bolong-bolong dan makan banyak korban #WaspadaPengusungKhilafahRI Happy Valentine MALUHEHE Hukuman Mati Kuat Ma'ruf

B. *Website*

Proses analisis sentimen dibuat dalam bentuk website yang mempunyai sistem untuk mengetahui sentimen warga *twitter* dalam kasus pemilihan umum 2024. Pada sistem yang terdapat dalam website sendiri mencakup *import data csv*, *preprocessing*, analisis sentimen, *clustering* algoritma *k-means* dan hasil akhir. Hasil dari

proses *crawling data* akan berbentuk *csv file* yang terdiri dari *tweets* masyarakat. Data yang ada di *file csv* masih berbentuk data mentah maka harus dilakukannya *preprocessing* pada sistem. Setelah melakukan *preprocessing* maka sistem akan melakukan proses pelabelan. Setelah sudah melakukan pelabelan maka masuk ke dalam *clustering* menggunakan algoritma *k-means* untuk mengetahui *cluster* pada tiap *tweets*. *Cluster* yang telah dibuat maka akan diuji tingkat akurasi menggunakan metode *silhouette coefficient*. Hasil akhir dari data berupa diagram bar, *cluster*, teks dan tabel yang menggunakan *html*, *css* dan *flask python* dengan *text editor Visual Studio Code*[8].

C. Preprocessing.

Setelah melakukan proses *crawling*, maka tahap selanjutnya yaitu tahapan *preprocessing*, data yang telah didapatkan masih tidak beraturan dan perlu dibersihkan agar tidak mengganggu jalannya proses analisis sentimen. Tahapan *preprocessing* terdiri dari *cleaning*, *case folding*, *stopword removal*, *tokenizing* dan *stemming*[6].

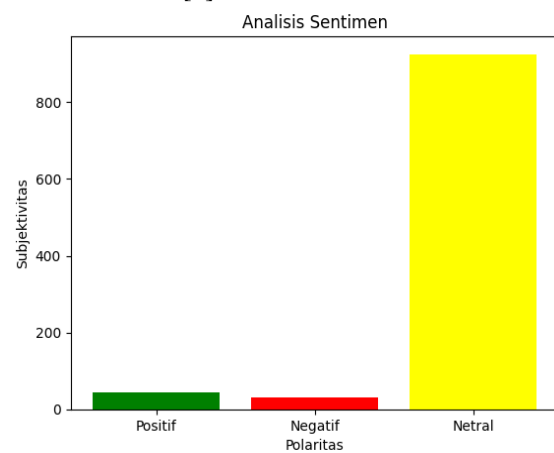
Tabel 2 Hasil Preprocessing

No	Tweets	Preprocessing
1.	Mari kita sukseskan Pemilu 2024 yang AMAN, DAMAI dan SEJUK di Kota Magelang	mari sukses pemilu 2024 aman damai sejuk kota magelang
2.	Semoga Pemilu 2024 berjalan lancar dan Jurdil. Tak ada lagi kampanye hitam, tak ada lagi politik identitas, tak ada lagi provokator. Sehingga dapat terwujudkan Pesta demokrasi yang berkualitas.	semoga pemilu 2024 berjalan lancar jurdil kampanye hitam politik identitas provokator terwujud pesta demokrasi kualitas
3.	Dengan penuh integritas, Kepala Kantor Imigrasi Kelas I TPI Gorontalo memimpin pembacaan ikrar diikuti oleh seluruh pegawai, berisi empat poin penting pernyataan netralitas dalam menyukseskan Pemilu 2024.	penuh integritas kepala kantor imigrasi kelas I gorontalo mimpin baca ikrar ikuti oleh pegawai isi poin nyata netralitas sukses pemilu 2024
4.	Stop politik identitas! Pemimpin itu punya visi bukan memprovokasi #pemilu2024	politik identitas pemimpin visi provorkasi
5.	Ini dia calon pemimpi di pemilu 2024 yang suka bikin jalanan bolong-bolong dan makan banyak korban #WaspadaPengusungK hilafahRI Happy	Calon mimpin pemilu 2024 jalan bolong bolong makan banyak korban

No	Tweets	Preprocessing
	Valentine MALUHEHE Hukuman Mati Kuat Ma'ruf	

D. Pelabelan

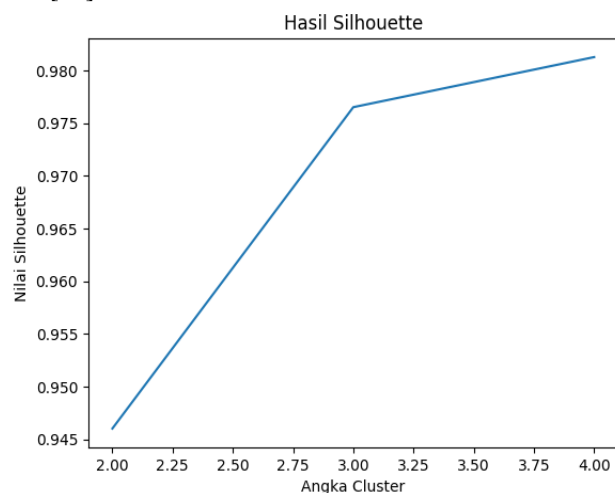
Pelabelan digunakan untuk mengetahui masing-masing opini masyarakat. Pelabelan dikategorikan menjadi 3 yang terdiri dari positif, netral dan negatif. Untuk mendapatkan hasil pelabelan maka perlu dilakukan perhitungan menggunakan polaritas dan subjektivitas. Hasil akhir dalam proses pelabelan dibuat dalam bentuk diagram bar. Sentimen positif memiliki persentase nilai sebesar 4.30%, sentimen netral memiliki persentase nilai sebesar 92.50% dan sentimen negatif memiliki persentase nilai sebesar 3.20% [9].



Gambar 1. Hasil Analisis

E. Silhouette Coefficient

Metode *silhouette Coefficient* digunakan untuk mengetahui tingkat ketepatan dari algoritma *K-Means Clustering*. Proses perhitungan ini akan menghitung setiap jarak dari semua titik yang ada dalam *cluster*. Proses perhitungan ini dilakukan dengan memakai *python*. Dari setiap *cluster* yang di iterasikan nilai *silhouette coefficient* semakin bertambah menuju angka 1, itu membuktikan bahwa hasil *clustering* menggunakan algoritma *k-means* memiliki hasil susunan yang kuat [10].



Gambar 2 Hasil Pengujian Silhouette Coefficient

IV. KESIMPULAN

Sesudah melakukan proses analisis sentimen diperoleh kesimpulan yang terdiri dari:

1. Berdasarkan hasil penelitian analisis sentimen didapati sebanyak 92.50% bersifat netral, 4.30% bersifat positif dan 3.20% bersifat negatif, jadi mayoritas pengguna media sosial twitter lebih banyak membuat respon netral pada kata kunci” pemilu 2024”
 2. Dalam penelitian ini tingkat akurasi yang digunakan untuk mengukur ketepatan analisis sentimen yang dibuat, nilai akurasi yang didapatkan menggunakan silhouette coefficient sebesar 0.981263 yang memiliki hasil susunan struktur yang kuat.
 3. Dalam penelitian analisis sentimen ini didapati bahwa kata yang sering muncul dalam penelitian ini yaitu adalah kata “pemilu”, “pemilihan,” prabowo,” data” dan” pilih”.
- Lalu untuk saran yang bisa diberikan untuk para peneliti selanjutnya untuk memperbaiki hasil dari analisis sentimen adalah:
- a. Memakai metode ataupun algoritma yang berbeda sebagai pembanding.
 - b. Melakukan clustering secara real time agar hasilnya selalu baru.
 - c. Menggunakan metode pengujian yang berbeda sehingga dapat mengetahui tingkat akurasi dengan menggunakan metode pengujian yang berbeda.
 - d. Memakai media sosial yang berbeda untuk mengetahui nilai sentimen masyarakat terhadap kasus yang sama.

DAFTAR PUSTAKA

- [1] P. P. O. Mahawardana, I. A. P. F. Imawati, and I. W. Dika, “Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap ‘Figure Pemimpin’ Menggunakan Python,” *J. Manaj. dan Teknol. Inf.*, vol. 12, no. 2, pp. 50–56, 2022, [Online]. Available: <https://ojs.mahadewa.ac.id/index.php/jmti/article/view/2111>
- [2] M. R. Fais Sya’ bani, U. Enri, and T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naive Bayes,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [3] R. Asmara, M. F. Ardiansyah, and M. Anshori, “Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier,” *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 193, 2020, doi: 10.35314/isi.v5i2.1095.
- [4] F. Rizqi Irawan, “Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearset Neighbors,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.v5i1.4267.
- [5] S. I. Safitri, C. Suhery, and S. Bahri, “Implementasi Algoritma K-Means Untuk Clustering Sentimen Pada Opini Kualitas Pelayanan Jasa Penerbangan,” *Coding J. Komput. dan Apl.*, vol. 09, no. 02, pp. 186–197, 2021, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jcskommipa/article/view/47377>
- [6] M. Hudha, E. Supriyati, and T. Listyorini, “Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naive Bayes Classifier,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
- [7] G. Sanko, S. Wasino, and T. Sutrisno, “Klasterisasi Sentimen Ulasan Pengguna Aplikasi Bca Mobile Pada Platform Google Play Store Dengan Algoritma K-Means Clustering”.
- [8] P. S. Zalukhu, T. Handhayani, M. Sitorus, T. Informatika, and U. Tarumanagara, “Kenaikan Bbm Di Indonesia Pada Media Sosial Twitter Menggunakan Metode Naive Bayes,” vol. 8, no. 1, 2023.
- [9] A. S. Amirul Haj, V. Amrizal, and A. Arini, “Analisis Sentimen Kinerja KPU Pemilu 2019 Menggunakan Algoritma K-Means Dengan Algoritma Confix Stripping Stemmer,” *J. Innov. Inf. Technol. Appl.*, vol. 2, no. 01, pp. 9–18, 2020, doi: 10.35970/jinita.v2i01.119.
- [10] M. A. Z. Larasati, N. A. S. Winarsih, M. S. Rohman, and G. W. Saraswati, “Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter,” *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 201, 2022, doi: 10.35889/progresif.v18i2.877.