

Analisis Sentimen Twitter Pengaruh Tokoh Politik dengan Menggunakan Metode K-Nearest Neighbor

I Made Surya Adi Palguna^{a1}, Ngurah Agus Sanjaya ER^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana, Bali

Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 80361, Bali, Indonesia

¹surya.adi.palguna067@student.unud.ac.id

²agus_sanjaya@unud.ac.id (Corresponding Author)

Abstract

Public opinion towards political figures can consist of positive and negative sentiments. Besides that, social media has developed which can be used as a forum for public opinion, one of which is Twitter. From this public opinion, sentiment analysis is formed which uses a classification algorithm. This work leverages the K-Nearest Neighbor (KNN) algorithm, which classifies data based on its similarity to existing data points. Tweets undergo preprocessing, followed by TF-IDF weighting for keyword importance and confusion matrix calculations for calculate the evaluation of algorithm. By analyzing the nearest neighbors, sentiment values are assigned. The KNN model achieved an accuracy of 84,06% for k = 5, precision of 86,70% for k = 5, recall of 95,89% for k = 7, and F1-score of 90,93% for k = 5, demonstrating its effectiveness in assessing sentiment and influence through Twitter data. This research contributes to the field of political communication by offering a robust method for analyzing public opinion and gauging the influence of political figures on social media platforms.

Keywords: Sentiment Analysis, K-Nearest Neighbor, TF-IDF, Confusion Matrix

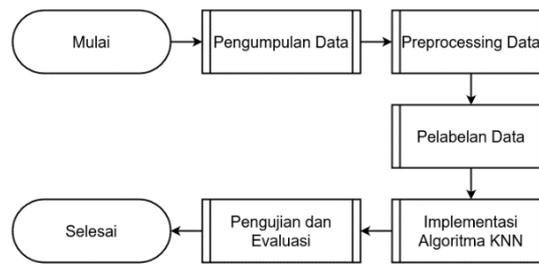
1. Pendahuluan

Pemilihan umum adalah proses demokratis di mana warga negara suatu negara secara langsung memilih wakil rakyat atau pejabat pemerintahan dalam menentukan pemimpin dan kebijakan negara. Adapun selama berlangsungnya pemilihan umum ini akan membawa dampak yang beragam antar tokoh politik di Indonesia terutama pengaruhnya. Sehingga diperlukannya opini masyarakat terhadap pengaruh tokoh politik. Opini masyarakat terhadap tokoh politik dapat berupa sentimen positif dan negatif. Disamping itu sudah berkembangnya media sosial yang bisa dijadikan wadah opini publik, salah satunya adalah Twitter. Dari opini publik tersebut, terbentuknya analisis sentimen yang dimana menggunakan suatu algoritma klasifikasi.

Terdapat beberapa algoritma yang bisa digunakan dalam klasifikasi data salah satunya K-Nearest Neighbor. K-Nearest Neighbors (KNN) merupakan pengembangan dari Nearest Neighbor (NN), di mana suatu pertanyaan diberi label kelas berdasarkan mayoritas suara dari k tetangga terdekatnya dalam himpunan pelatihan, menjadikannya salah satu metode klasifikasi sederhana dalam pengenalan pola [1]. Beberapa penelitian sebelumnya sudah menggunakan algoritma tersebut dengan Akurasi sebesar 67.2% sampai 89.4% pada nilai k = 3, k = 5, k = 7, k = 8, dan k = 13, Rasio Error sebesar 10.6% pada nilai k = 3, Recall sebesar 61.56% sampai 78.24% pada nilai k = 7 dan k = 15, Presisi sebesar 56.94% sampai 92.38% pada nilai k = 5 dan k = 7, serta AUC sebesar 0.894 sampai 0.916 pada nilai k = 7 dan k = 8 [2] [3] [4] [5] [6].

2. Metode Penelitian

Dalam melakukan proses analisis sentimen ini akan dilakukan tahapan penelitian sebagai berikut.



Gambar 1. Alur Metode Penelitian Analisis Sentimen

2.1. Pengumpulan Data

Pada penelitian analisis sentimen ini menggunakan data sebanyak 2000 tweet dari API Twitter dengan kata kunci “anies”, “prabowo”, “ganjar”, “muhammad”, “gibran”, dan “mahfud” menggunakan library Python yang bernama tweet-harvest. Setelah didapat data tersebut akan dibagi menjadi data latih dan data validasi dengan perbandingan 80:20 sehingga data latih yang digunakan sebanyak 1600 data dan data validasi sebanyak 400 data.

2.2. Preprocessing Data

Tahapan preprocessing data bertujuan untuk mempersiapkan dataset agar dapat diolah oleh algoritma analisis data. Proses ini dapat dibagi menjadi beberapa tahap, yaitu (1) Case Folding, (2) Tokenizing, (3) Stopword Removal, dan (4) Stemming.

- a. Case Folding
Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) atau huruf besar (uppercase). Selain itu adapun proses lainnya seperti membersihkan data duplicate, data missing value, dan cleaning data tweet.
- b. Tokenizing
Tokenisasi adalah proses memecah teks menjadi bagian-bagian kecil yang disebut token. Pada proses ini akan menggunakan function yang ada di bahasa pemrograman Python.
- c. Stopword Removal
Stopwords adalah kata-kata umum yang sering muncul dalam suatu bahasa tetapi kurang memberikan makna signifikan dalam analisis teks. Pada proses ini akan menghilangkan stopwords dengan menggunakan library Python yang bernama Sastrawi.
- d. Stemming
Stemming adalah proses menghilangkan afiks dari kata untuk menghasilkan bentuk dasarnya, yang disebut stem. Pada proses stemming akan menggunakan library Python yang bernama Sastrawi.

2.3. Pelabelan Data

Tahapan pelabelan data bertujuan untuk menetapkan label atau kategori pada setiap instan data. Pelabelan data dalam analisis sentimen Twitter dengan metode K-Nearest Neighbor berfungsi untuk mengkategorikan tweet ke dalam kelas sentimen yang tepat, memungkinkan model memahami pola sentimen dan meningkatkan akurasi prediksi.

2.4. Implementasi Algoritma KNN

Tahapan implementasi algoritma KNN bertujuan untuk mengimplementasikan algoritma klasifikasi KNN (K-Nearest Neighbors) pada dataset. Pada pengimplementasian menggunakan kedua algoritma tersebut akan menggunakan parameter berupa matriks tokenisasi semua data

latih, label semua data latih, matriks tokenisasi yang akan dihasilkan labelnya, dan nilai k. Adapun nilai k yang digunakan dalam penelitian ini sebesar 3, 5, dan 7. Sehingga keluaran yang dihasilkan berupa label dari matriks tokenisasi tersebut. Adapun menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) yang berfungsi untuk mengevaluasi seberapa penting suatu kata dalam sebuah dokumen dengan memperhitungkan frekuensinya dalam dokumen tersebut dan invers proporsional dengan frekuensi kemunculannya di seluruh koleksi dokumen.

2.5. Pengujian dan Evaluasi

Tahapan evaluasi bertujuan untuk menguji dan mengukur kinerja algoritma klasifikasi yang telah diimplementasikan. Pada penelitian ini akan menggunakan data validasi sebesar 400 data dan evaluasi yang diukur dalam penelitian ini meliputi Akurasi, Presisi, Recall, dan F1-Score dengan menggunakan confusion matrix.

3. Hasil dan Pembahasan

3.1. Pembahasan

Pada penelitian ini menggunakan Google Colab dengan beberapa library Python. Adapun proses dari metode dalam penelitian ini dimulai dari pengumpulan data sebanyak 2000 data yang dapat digambarkan hasilnya sebagai berikut.

	created_at	full_text	username
0	Mon Jan 08 02:10:09 +0000 2024	hilih dulu partai Mercedes pas dukung Anies pa...	arthrw_
1	Mon Jan 08 02:10:09 +0000 2024	@ZulkifilLubis69 @prabowo Y Allah jadikan Anie...	Rawing1274
2	Mon Jan 08 02:10:08 +0000 2024	@Malika6027 Anies alumni UGM Ganjar alumni UGM...	penjagahatii009
3	Mon Jan 08 02:10:03 +0000 2024	Masyarakat bisa dianghap memneri amal jaryah ...	wigRahman
4	Mon Jan 08 02:10:01 +0000 2024	@mihrabku @aniesbaswedan @prabowo @ganjarprano...	PietSofyatuddin
...
2250	Mon Jan 08 03:15:14 +0000 2024	Pake #GanjarMahfudRebound itu seru banget, tem...	bawerman68390
2251	Mon Jan 08 03:15:14 +0000 2024	Pake #JNKBersamaGanjarMahfud itu seru banget, ...	mitchell_j70489
2252	Mon Jan 08 03:15:14 +0000 2024	Pake #GanjarMahfudRebound itu seru banget, tem...	MitchellCh36585
2253	Mon Jan 08 03:15:14 +0000 2024	Dalam era #GanjarMahfudRebound, Ganjar Pranowo...	AnthonyFis18905
2254	Mon Jan 08 03:15:13 +0000 2024	Yudhalmamf Twillight_Loid Ke Mana Aje Tw...	NatalieGrace3

2255 rows × 3 columns

Gambar 2. Pengumpulan Data

Kemudian adapun preprocessing data yang terbagi menjadi Case Folding, Tokenizing, Stopword Removal, dan Stemming yang dapat digambarkan hasilnya sebagai berikut.

	case_folding	stopword_removal	stemming
0	hilih dulu partai mercedes pas dukung anies pa...	[hilih, dulu, partai, mercedes, pas, dukung, p...	hilih dulu partai mercedes pas dukung pake pin...
1	y allah jadikan anies baswedan sbg presiden ri...	[y, allah, jadikan, baswedan, sbg, presiden, r...	y allah jadi baswedan sbg presiden rbiar urus...
2	anies alumni ugm ganjar alumni ugm prabowo alu...	[alumni, ugm, alumni, ugm, alumni, capres, 2014]	alumni ugm alumni ugm alumni capres 2014
3	masyarakat bisa dianghap memneri amal jaryah ...	[masyarakat, dianghap, memneri, amal, jaryah,...	masyarakat dianghap memneri amal jaryah rp 45...
4	saya pikir tadi malam mereka memang berbagi tu...	[pikir, tadi, malam, memang, berbagi, tugas, m...	pikir tadi malam memang bagi tugas mantik kubu...
...
2242	pasangan capres cawapres ganjar pranowo mahfud...	[pasangan, capres, cawapres, pranowo, md, calo...	pasang capres cawapres pranowo md calon pimpin...
2243	ganjar pranowo dan mahtud md mendukung dengan ...	[pranowo, mahtud, md, mendukung, tegas, upaya,...	pranowo mahtud md dukung tegas upaya cipta leg...
2245	pasangan capres dan cawapres ri nomor 03 ganja...	[pasangan, capres, cawapres, ri, nomor, 03, pr...	pasang capres cawapres ri nomor 03 pranowo md ...
2246	dalam era ganjar pranowo dan mahtud md memimpi...	[era, pranowo, mahtud, md, memimpin, teladan, ...	era pranowo mahtud md pimpin teladan ajak masy...
2254	yudhaimamf twillightoid ke mana aje twillight...	[yudhaimamf, twillightoid, mana, aje, twillig...	yudhaimamf twillightoid mana aje twillightoi...

2040 rows × 3 columns

Gambar 3. Preprocessing Data

Setelah itu akan dilakukan pelabelan data setelah melakukan preprocessing data sehingga dapat digambarkan hasilnya sebagai berikut.

	preprocessing	tweet_english	sentimen
0	hilih dulu partai mercedes pas dukung pake pin...	i first chose the mercedes party when supporti...	Positif
1	y allah jadi baswedan sbg presiden ribiar urus...	o god, so baswedan as president of the republi...	Negatif
2	alumni ugm alumni ugm alumni capres 2014	ugm alumni alumni ugm alumni candidate 2014	Positif
3	masyarakat dianghap memneri amal jariah rp 45...	the people are arrested by charity jariah rp...	Positif
4	pikir tadi malam memang bagi tugas mantik kubu...	thinking last night indeed for the task of the...	Positif
...
2035	pasang capres cawapres pranowo md calon pimpin...	install the vice presidential candidate pranow...	Positif
2036	pranowo mahfud md dukung tegas upaya cipta leg...	pranowo mahfud md firmly supports efforts to c...	Negatif
2037	pasang capres cawapres ri nomor 03 pranowo md ...	install the indonesian vice presidential candi...	Positif
2038	era pranowo mahfud md pimpin teladan ajak masy...	era pranowo mahfud md lead an example inviting...	Negatif
2039	yudhaimamf twillightloid mana aje twillightloi...	yudhaimamf twillightloid where aje twillightlo...	Positif

2040 rows × 3 columns

Gambar 4. Pelabelan Data

Setelah itu akan dilakukan implementasi algoritma KNN dan pembobotan TF-IDF yang dimana data tersebut akan dibagi menjadi data latih dan data validasi dengan perbandingan 80:20 sehingga data latih yang digunakan sebanyak 1600 data dan data validasi sebanyak 400 data. Sehingga dapat digambarkan hasilnya sebagai berikut.

	tweet_data	label	matrix
0	hilih dulu partai mercedes pas dukung pake pin...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
1	y allah jadi baswedan sbg presiden ribiar urus...	0	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2	alumni ugm alumni ugm alumni capres 2014	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
3	masyarakat dianghap memneri amal jariah rp 45...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
4	pikir tadi malam memang bagi tugas mantik kubu...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
...
2035	pasang capres cawapres pranowo md calon pimpin...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2036	pranowo mahfud md dukung tegas upaya cipta leg...	0	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2037	pasang capres cawapres ri nomor 03 pranowo md ...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2038	era pranowo mahfud md pimpin teladan ajak masy...	0	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2039	yudhaimamf twillightloid mana aje twillightloi...	1	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...

2040 rows × 3 columns

Gambar 5. Implementasi Algoritma KNN

Hingga setelah proses tersebut akan dilakukan proses pengujian dan evaluasi menggunakan perhitungan confusion matrix baik untuk Akurasi, Presisi, Recall, dan F1-Score yang dimana memberikan gambaran jelas tentang kinerja dengan menyajikan informasi detil tentang hasil klasifikasi, termasuk jumlah prediksi benar dan salah untuk setiap kelas.

3.2 Hasil

Berikut adalah hasil analisis sentimen Twitter terhadap pengaruh tokoh politik dengan menggunakan metode K-Nearest Neighbor:

Tabel 1. Hasil Analisis Sentimen

Nilai k	Akurasi	Presisi	Recall	F1-Score
3	83,08%	85,78%	95,60%	90,43%
5	84,06%	86,70%	95,60%	90,93%
7	82,59%	85,15%	95,89%	90,20%

Dari tabel di atas, dapat dilihat bahwa nilai akurasi, presisi, dan F1-score tertinggi diperoleh pada nilai k = 5. Hal ini menunjukkan bahwa metode K-Nearest Neighbor dengan nilai k = 5 dapat memberikan hasil analisis sentimen yang paling akurat.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa penelitian tentang analisis sentimen pengguna Twitter terhadap pengaruh tokoh politik dengan menggunakan metode K-Nearest Neighbor telah berhasil dilakukan. Dengan menggunakan confusion matrix, hasil akurasi yang terbesar 84,06% pada k = 5, nilai presisi terbesar 86,70% pada k = 5, recall terbesar 95,89% pada k = 7, dan F1-score terbesar 90,93% pada k = 5. Hasil penelitian ini menunjukkan bahwa metode K-Nearest Neighbor dapat digunakan untuk menganalisis sentimen Twitter terhadap pengaruh tokoh politik dengan akurasi yang cukup tinggi. Hal ini menunjukkan bahwa analisis sentimen Twitter dapat menjadi alat yang berguna untuk mengukur pengaruh tokoh politik terhadap masyarakat.

Daftar Pustaka

- [1] J. Gou, L. Du, Y. Zhang and T. Xiong, "A New Distance-Weighted K-Nearest Neighbor Classifier," *Journal of Information & Computational Science*, pp. 1429-1436, 2012.
- [2] M. S. Alrajak, I. Ernawati e I. Nurlaili, "Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K-Nearest Neighbor (K-NN)," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 110-122, 2020.
- [3] A. Asro'i e H. Februariyanti, "Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor," *Jurnal Khatulistiwa Informatika*, vol. X, n^o 1, pp. 17-24, 2022.
- [4] A. Deviyanto e M. D. R. Wahyudi, "Penerapan Analisis Sentimen pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. III, n^o 1, pp. 1-13, 2018.
- [5] S. Ernawati e R. Wati, "Penerapan Algoritma K-Nearest Neighbors pada Analisis Sentimen Review Agen Travel," *Jurnal Khatulistiwa Informatika*, vol. VI, n^o 1, pp. 64-69, 2018.
- [6] R. Sari, "Analisis Sentimen pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *Jurnal Sains dan Manajemen*, vol. VIII, n^o 1, pp. 10-17, 2020.

Halaman ini sengaja dibiarkan kosong