
Implementasi Text Clustering Terkait Pilpres 2024 Menggunakan Metode *K-Means*

Miquel Yosafat^{1*}, Jatmika²

Program Studi Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel Yogyakarta
Jln. Ukrim No.KM. 11, Kadirojo I, Purwomartani, Kalasan, Sleman, Daerah Istimewa Yogyakarta 55571, Indonesia

E-mail: miquel.y19@student.ukrimuniversity.ac.id¹

Abstrak

Penelitian ini fokus pada implementasi metode K-Means Clustering untuk menganalisis opini masyarakat terhadap Pilpres 2024. Algoritma K-Means, sebagai metode data mining tanpa arahan, digunakan untuk mengelompokkan data opini yang memiliki karakteristik serupa. Hasil analisis cluster memastikan ketiadaan teks yang bersifat sara atau sarkasme dalam data Twitter yang diambil. Kluster dibagi dan dikategorikan berdasarkan pendekatan teks, dan hasilnya menunjukkan bahwa kata "pilpres" muncul paling banyak dengan total count 1778, sementara kata "ahy" muncul paling sedikit dengan total count 15. Penelitian ini memberikan wawasan mendalam terkait persepsi masyarakat terhadap Pilpres 2024 melalui analisis cluster data opini di media sosial.

Abstract

This research focuses on implementing the K-Means Clustering method to analyze public opinion regarding the 2024 presidential election. The K-Means algorithm, as a data mining method without direction, is used to group opinion data that has similar characteristics. The results of the cluster analysis confirmed the absence of text that was sarcastic or sarkastic in the Twitter data taken. Clusters were divided and categorized based on the text approach, and the results showed that the word "pilpres" appeared the most with a total count of 1778, while the word "ahy" appeared the least with a total count of 15. This research provides in-depth insight into public perceptions of the 2024 presidential election through analysis opinion data clusters on social media.

Info Naskah:

Naskah masuk : 04 Januari 2024

Direvisi: 09 Januari 2024

Diterima : 10 Januari 2024

Keywords:

Presidential Election;
Clustering;
K-Means;
Presidential Candidate;
Twitter

*Penulis korespondensi:

Miquel Yosafat

E-mail: miquel.y19@student.ukrimuniversity.ac.id

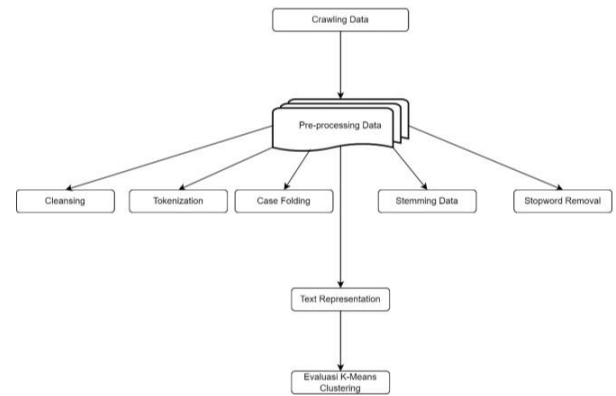
1. Pendahuluan

Pemilu merupakan satu-satunya prosedur demokrasi yang melegitimasi kewenangan dan tindakan para wakil rakyat untuk melakukan tindakan tertentu. Pemilu adalah mekanisme sirkulasi dan regenerasi kekuasaan. Pemilu juga satu-satunya cara untuk menggantikan kekuasaan lama tanpa melalui kekerasan dan kudeta. Melalui pemilu rakyat dapat menentukan sikap politiknya untuk tetap percaya pada pemerintah lama, atau menggantikannya dengan yang baru. Dengan kata lain, pemilu merupakan sarana penting dalam mempromosikan dan meminta akuntabilitas dari para pejabat publik salah satu nya pemilihan umum presiden yang diselenggarakan setiap lima tahun sekali. (Fitria Rini, 2015). Media sosial merupakan salah satu sarana yang digunakan oleh pengguna untuk mengakses, berbagi, dan berdiskusi tentang opini-opini masyarakat terkait dengan politik. Salah satunya adalah Twitter, Twitter adalah salah satu media sosial yang populer saat ini. Dengan menggunakan metode *Text Clustering* di Twitter, kita dapat memperoleh pemahaman tentang bagaimana opini-opini masyarakat terhadap Pilpres 2024 (Fridom Mailo et al., 2019).

Dalam penelitian ini akan dilakukan *clustering* terkait pilpres 2024 yang akan diungkap melalui media sosial yaitu Twitter. Algoritma yang diterapkan pada penelitian ini adalah algoritma *K-Means Clustering*. Algoritma *K-Means Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*) dan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. Pengelompokan terbagi menjadi dua, yaitu komplet dan parsial. Jika semua data dapat bergabung menjadi satu, dapat dikatakan semua data kompak menjadi satu kelompok. Pada clustering ini terdapat beberapa algoritma pengelompokan untuk mengelompokkan data secara mudah. Salah satunya adalah algoritma *K-Means* yang merupakan metode analisis kelompok yang mengarah pada partisi N objek pengamatan ke dalam K kelompok, di mana setiap objek pengamatan sebuah kelompok data dengan mean (rata-rata) terdekat.

2. Metode Penelitian

Metode untuk penelitian yang akan digunakan pada penelitian implementasi *text clustering* dengan menggunakan Bahasa pemrograman python. Penelitian ini memiliki beberapa Langkah-langkah yaitu crawling data melalui media sosial twitter, *pre-processing* data yang meliputi 5 langkah (*Cleansing, Tokenization, Case Folding, Stemming Data, dan Stopword Removal*), *Text Representation*, dan Evaluasi *K-Means*.



Gambar 2.1 Langkah Metode *Text Clustering*

3. Hasil dan Pembahasan

3.1 Crawling Data

Crawling data menggunakan harvest dan mengambil data dari media sosial Twitter dengan kata kunci Pilpres 2024 rentang waktu 1 Januari 2022 hingga 1 Mei 2023 dan mendapatkan hasil tweet data sebanyak 4.155 data Tweet. Hasil data crawling bisa dilihat pada gambar 3.1.

Gambar 3.1 Crawling Data

3.2 Pre-processing Data

Proses *Preprocessing* data akan menyeleksi beberapa data yang tidak digunakan dengan menerapkan beberapa langkah mulai dari *cleansing, tokenization, case folding, stopwords removal* dan *stemming* data. Setelah melakukan langkah-langkah *pre-processing* data maka hasil dari crawling data 4.155 tweet menjadi 1.816 tweet setelah melakukan semua langkah preprocessing data. Hasil dari *Preprocessing* data dapat dilihat pada Gambar 3.2.

Gambar 3.2 Pre-processing Data

3.3 Text Representation

Text Representatiton merupakan proses untuk mengubah teks menjadi vektor, untuk menggunakan *metode K-Means* data harus berupa vektor. Hasil dari *Text Representation* dapat dilihat pada Gambar 3.3.

```
Jumlah dokumen: 1816
Jumlah fitur: 6513

Representasi Vector:
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

Gambar 3.3 Text Representation

3.4 K-Means Clustering

Klasifikasi *K-Means Clustering* hanya mengelompokkan data ke (klaster) sesuai dengan data yang sudah dibagi menjadi 10 klaster. Hasil dari pembagian *cluster* dapat dilihat pada Gambar 3.4.

	Tweet	Cluster
0	bismillah ganjarpranowo utk ri	1
1	bismilah ganjar pranowo tugiman moga tdk presi...	2
2	kerozaok prabowo kalah pilpres rakyat indonesi...	7
3	ganjar pranowo satu partai demokrasi indonesia...	0
4	jakarta cnn indonesia ketua partai gerindra pr...	3
...
295	gus yaqut dapat ganjarerick pasang ideal pilpr...	9
296	gaet menang pilpres nnduo pasang ganjar pranow...	9
297	yurissasamosir anies jubir Jokowi gub DKI Pung...	7
298	banding pasang nganjar pranowo amp erick thohi...	5
299	geloraco pilpres klo curang makan korban rakya...	1

Gambar 3.4 K-Means Clustering

3.5 Hasil Cluster

3.5.1 Cluster ke-1

Cluster ke-1 ini berhubungan dengan topik Pilpres 2024 di Indonesia. *Cluster* ini mencakup kata-kata yang terkait dengan politik dan pemilihan presiden, seperti "pilpres", "presiden", dan "partai". Terdapat juga kata-kata yang menunjukkan adanya pembicaraan tentang koalisi politik, seperti "koalisi" dan "ppp" (Partai Persatuan Pembangunan). Selain itu, nama-nama tokoh politik juga muncul dalam kluster ini, seperti "ganjar" (Ganjar Pranowo), yang merupakan calon presiden potensial dalam Pilpres 2024. Kata-kata seperti "satu" dan "indonesia" menunjukkan adanya fokus pada persatuan dan identitas nasional dalam konteks Pilpres 2024.

	word	count
0	koalisi	87
1	satu	74
2	indonesia	74
3	pilpres	66
4	partai	53
5	kib	42
6	presiden	34
7	ppp	33
8	ganjar	31
9	pranowo	25

Gambar 3.5 Cluster ke-1

3.5.2 Cluster ke-2

Cluster ke-2 ini mencakup kata-kata yang terkait dengan Pilpres 2024, seperti "pilpres", "indonesia", "jokowi" (Joko Widodo), dan "presiden". Terdapat juga kata-kata yang menunjukkan adanya dukungan dan partisipasi politik dari rakyat, seperti "rakyat", "pilih", dan "dukung". *Kluster* ini juga mencakup kata-kata yang berkaitan dengan partai politik, menunjukkan adanya peran partai dalam konteks Pilpres 2024.

	word	count
0	pilpres	480
1	indonesia	473
2	jokowi	123
3	rakyat	115
4	presiden	107
5	pilih	103
6	politik	101
7	dukung	95
8	partai	78
9	menang	72

Gambar 3.6 Cluster ke-2

3.5.3 Cluster ke-3

Cluster ke-3 ini mencakup kata-kata yang terkait dengan salah satu bacapres, yaitu "ganjar" (Ganjar Pranowo) dan "pranowo". Hal ini menunjukkan adanya perbincangan atau perhatian terhadap Ganjar Pranowo sebagai calon dalam Pilpres 2024. Terdapat kata-kata yang menunjukkan adanya dukungan terhadap Ganjar Pranowo, seperti "dukung" dan "menang". Hal ini mengindikasikan adanya harapan atau aspirasi bahwa Ganjar Pranowo akan berhasil dalam Pilpres 2024. Kata-kata seperti "pilpres", "indonesia", "presiden", dan "calon" menunjukkan fokus pada konteks Pilpres 2024 dan pemilihan presiden di Indonesia.

	word	count
0	ganjar	206
1	pilpres	143
2	indonesia	134
3	pranowo	132
4	presiden	70
5	dukung	61
6	menang	36
7	ganjarpranowo	34
8	pimpin	33
9	calon	32

Gambar 3.7 Cluster ke-3

3.5.4 Cluster ke-4

Cluster ke-4 ini mencakup kata-kata yang terkait dengan tokoh politik Prabowo Subianto, seperti "prabowo" dan "subianto". Hal ini menunjukkan adanya perbincangan atau perhatian terhadap Prabowo Subianto dalam konteks Pilpres 2024. Terdapat kata-kata yang menunjukkan adanya dukungan terhadap Prabowo Subianto, seperti "dukung" dan "ikutpakde". Hal ini mengindikasikan adanya harapan atau aspirasi bahwa Prabowo Subianto akan terlibat atau berperan penting dalam Pilpres 2024. Kata-kata seperti "pilpres", "indonesia", "presiden", "gerindra", dan "partai" menunjukkan fokus pada konteks Pilpres 2024 dan peran Gerindra sebagai partai politik yang terkait dengan Prabowo Subianto.

	word	count
0	prabowo	175
1	subianto	159
2	indonesia	134
3	pilpres	131
4	presiden	66
5	gerindra	60
6	partai	58
7	ikutpakde	55
8	ketua	38
9	dukung	31

Gambar 3.8 Cluster ke-4

3.5.5 Cluster ke-5

Cluster ke-5 ini mencakup kata-kata yang terkait dengan tokoh politik Anies Baswedan, seperti "anies" dan "baswedan". Hal ini menunjukkan adanya perhatian terhadap Anies Baswedan sebagai salah satu tokoh yang mungkin terlibat dalam Pilpres 2024. Terdapat kata-kata yang menunjukkan adanya dukungan terhadap Anies Baswedan, seperti "dukung" dan "aniesbaswedan". Hal ini mengindikasikan adanya harapan atau aspirasi bahwa Anies Baswedan akan mendapatkan dukungan dalam Pilpres 2024. Selain itu, kata-kata seperti "pilpres", "indonesia",

Gambar 3.10 Cluster ke-6

3.5.7 Cluster ke-7

Cluster ke-7 Kluster ini mencakup kata-kata yang terkait dengan Pilpres, seperti "pilpres", "pimpin", dan "ubah". Hal

"presiden", "pilih", dan "ubah" menunjukkan fokus pada konteks Pilpres 2024 dan potensi perubahan yang diinginkan oleh para pendukung Anies Baswedan.

	word	count
0	anies	220
1	pilpres	152
2	indonesia	140
3	baswedan	79
4	dukung	50
5	aniesbaswedan	48
6	menang	48
7	presiden	44
8	pilih	37
9	ubah	34

Gambar 3.9 Cluster ke-5

3.5.6 Cluster ke-6

Cluster ke-6 ini mencakup kata-kata yang terkait dengan survei, seperti "survei", "hasil", dan "lembaga". Hal ini menunjukkan adanya perbincangan tentang hasil survei terkait dengan Pilpres di Indonesia. Kata-kata ini juga menunjukkan perhatian terhadap lembaga-lembaga yang melakukan survei terkait Pilpres. Selain itu, kata-kata seperti "pilpres", "indonesia", "presiden", "capres", dan "subianto" menunjukkan fokus pada konteks Pilpres di Indonesia dan keterlibatan tokoh politik Prabowo Subianto. Prabowo Subianto disebutkan sebagai tokoh yang terkait dengan hasil survei yang dibahas dalam kluster ini.

	word	count
0	survei	113
1	pilpres	82
2	indonesia	74
3	hasil	64
4	lembaga	47
5	prabowo	34
6	presiden	31
7	capres	28
8	thohir	26
9	subianto	26

ini menunjukkan adanya perbincangan dan perhatian terhadap kepemimpinan dan perubahan terkait dengan Pilpres. Terdapat kata-kata yang mengacu pada tokoh Anies Baswedan dan AHY (Agus Harimurti Yudhoyono), seperti "aniesahy", "indonesianahy", dan "ahy". Hal ini menunjukkan adanya keterkaitan antara kluster ini dengan kedua tokoh tersebut. Juga terdapat kata-kata yang mengacu pada partai-partai politik terkait, seperti "jatengspn" (Jateng Solidaritas Partai Nasional), "ndemocratsnpidato" (Partai Demokrat), dan "indonesianpatidemokrat" (Partai Demokrat Indonesia). Hal ini menunjukkan adanya keterkaitan dengan partai-partai yang terkait dengan Anies Baswedan dan AHY.

	word	count
0	pilpres	47
1	jatengspn	47
2	ahyn	43
3	ndemocratsnpidato	41
4	aniesahy	33
5	indonesianpatidemokrat	24
6	indonesianahy	23
7	pimpin	23
8	ubah	23
9	ahy	15

Gambar 3.11 Cluster ke-7

3.5.7 Cluster ke-8

Cluster ke-8 terdapat perbincangan terkait Pilpres 2024 di Indonesia. Beberapa kata kunci yang muncul dalam kluster ini adalah "pilpres", "Indonesia", "Prabowo", "maju", "dukung", "Puan", "menang",

	word	count
0	pilpres	363
1	indonesia	298
2	prabowo	141
3	maju	70
4	dukung	65
5	puan	60
6	menang	50
7	ganjar	45
8	presiden	43
9	capres	33

Gambar 3.12 Cluster ke-8

3.5.9 Cluster ke-9

Cluster ke-9 ini mencakup kata-kata yang terkait dengan Pilpres 2024, seperti "pilpres", "indonesia", "erick" (Erick Thohir), dan "thohir". Terdapat juga kata-kata yang menunjukkan adanya peran Erick Thohir sebagai menteri dan kepemimpinan di sektor

	word	count
0	pilpres	208
1	erick	198
2	indonesia	190
3	thohir	175
4	menteri	128
5	bumn	122
6	pimpin	100
7	erickthohir	92
8	cawapres	87
9	dukung	77

Gambar 3.13 Cluster ke-9

3.5.10 Cluster ke-10

Cluster ke-10 ini mencakup kata-kata yang terkait dengan Pilpres 2024, seperti "pilpres", "indonesia", "ganjar" (Ganjar Pranowo), dan "thohir" (Erick Thohir). Terdapat juga kata-kata yang menunjukkan adanya pembahasan tentang pasangan atau duet dalam konteks Pilpres 2024, seperti "pasang", "duet", dan "erickthohir" (Erick Thohir sebagai calon presiden).

	word	count
0	pilpres	106
1	indonesia	93
2	pasang	91
3	ganjar	73
4	duet	62
5	erick	55
6	pranowo	46
7	thohir	43
8	erickthohir	37
9	capres	31

Gambar 3.14 Cluster ke-10

BUMN (Badan Usaha Milik Negara). Kluster ini juga mencakup kata-kata yang menunjukkan adanya dukungan terhadap Erick Thohir dan kemungkinan peranannya dalam Pilpres 2024.

4. Kesimpulan

Penelitian ini mengambil data dari Twitter dengan kata kunci Pilpres 2024 sebanyak 4.155 tweet dan setelah dilakukan *preprocessing* data menjadi 1.816 tweet. Melalui analisis *cluster*, dapat dievaluasi tidak ada teks yang berbau sara atau sarkasme sesuai dengan membagi klaster dan mengelompokkan klaster sesuai dengan pendekatan teks- teks dalam data Twitter. Berdasarkan hasil proses dari rata-rata tweet pada setiap klaster yang paling banyak kata yang muncul adalah pilpres dengan total *count* 1778 dan kata yang paling sedikit adalah ahy dengan total *count* 15.

Daftar Pustaka

- [1] Adhe, D., Rachman, C., Goejantoro, R., Deny, F., & Amijaya, T. (2020). Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering. *Jurnal EKSPONENSIAL*, 11.
- [2] Aulia, D., Safii, M., Suhendro, D., Studi, M. P., Informasi, S., Tunas Bangsa, S., & Tunas Bangsa, A. (n.d.). Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja ASN. 6, 47. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>.
- [3] Fitria Rini. (2015). BADAN PENGAWAS PEMILU DALAM SISTEM KETATANEGARAAN INDONESIA DALAM PERSPEKTIF KOMUNIKASI POLITIK.
- [4] Fridom Mailo, F., Lazuardi, L., Manajemen dan kebijakan Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan Universitas Gadjah Mada, K., Sistem Informasi Manajemen Kesehatan Fakultas Kedokteran, D., Masyarakat dan Keperawatan, K., & Gadjah Mada, U. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. In *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health* (Vol. 4, Issue 1).
- [5] Han, J., Kamber, M., Melton, J., Buxton, S., Teorey, T. J., Lightstone, S. S., Nadeau, T. P., Celko, J., Witten, I., Frank, E., Simson, G. C., Witt, G. C., Schiller, J., Voisard, A., Halpin, T., Evans, K., Hallock, P., Maclean, B., Ceri, S., ... Voisard, A. (n.d.). *Designing Data-Intensive Web Applications*.
- [6] Handoko Koko. (2016). PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI (STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN). 2.