



Jurnal Review Pendidikan dan Pengajaran
<http://journal.universitaspahlawan.ac.id/index.php/jrpp>
 Volume 7 Nomor 3, 2024
 P-2655-710X e-ISSN 2655-6022

Submitted : 29/05/2024
 Reviewed : 01/06/2024
 Accepted : 11/06/2024
 Published : 15/06/2024

Dwi Purbo Yuwono¹
Irwan Budi Santoso²
Ririen
Kusumawati³

PEMETAAN SENTIMEN MASYARAKAT TERHADAP PILPRES 2024 DENGAN ALGORITMA SELF-ORGANIZING MAP

Abstrak

Pemilihan Umum (PEMILU) adalah salah satu cara untuk memilih presiden, kepala daerah, dan anggota parlemen yang berlangsung setiap lima tahun sekali. Dalam memasuki tahun-tahun politik saat ini akan banyak bertebaran informasi dan komentar dari masyarakat terhadap pelaksanaan pemilu, komentar atau pendapat yang disampaikan akan sangat beragam dimulai dari dukungan terhadap pelaksanaan pemilu, penggiringan opini publik, ujaran kebencian dan komentar-komentar lainnya. Kemajuan teknologi saat ini mengakibatkan penyampaian pendapat dapat dengan mudah dipublikasikan melalui media sosial, salah satunya adalah melalui media twitter, twitter menjadi salah satu media sosial yang paling sering digunakan masyarakat dalam mengemukakan pendapatnya karena dianggap bebas. Oleh karena itu, pada penelitian ini diusulkan pemetaan sentimen atau opini masyarakat tentang Pilpres melalui X-Twitter, baik itu positif, negatif, atau netral dengan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan metode Self-Organizing Maps (SOM). Dari hasil penelitian didapatkan bahwa Algoritme TF-IDF dan Self-Organizing Maps (SOM) dengan sentimen cuitan pengguna Twitter dengan Hasil pengujian masing-masing model dengan menggunakan confusionmatrix didapatkan rata-rata accuracy sebesar 81%, precision 80,3%, recall 81%, dan f-measure 80%.

Kata Kunci: Pemilu, Twitter, TF-IDF, SOM.

Abstract

Pemilihan Umum (PEMILU) are one way to elect presidents, regional heads, and members of parliament which take place every five years. In entering the current political years, there will be a lot of information and comments from the public on the implementation of elections, comments or opinions conveyed will be very diverse starting from support for the implementation of elections, garnering public opinion, hate speech and other comments. Current technological advances result in the delivery of opinions can be easily published through social media, one of which is through Twitter media, Twitter is one of the social media most often used by people in expressing their opinions because it is considered free. Therefore, in this study, it is proposed to spy on public sentiment or opinion about the Presidential Election through X-Twitter, be it positive, negative, or neutral using Term Frequency-Inverse Document Frequency (TF-IDF) and Self-Organizing Maps (SOM) methods. From the results of the study, it was found that the TF-IDF Algorithm and Self-Organizing Maps (SOM) with the sentiment of Twitter users' tweets with the results of testing each model using confusionmatrix obtained an average accuracy of 81%, precision 80.3%, recall 81%, and f-measure 80%.

Keywords: Pemilu, Twitter, TF-IDF, SOM.

PENDAHULUAN

Pemilihan Umum (PEMILU) adalah salah satu cara untuk memilih presiden, kepala daerah, dan anggota parlemen yang berlangsung setiap lima tahun sekali. Pelaksanaan Pemilu dilakukan oleh penyelenggara pemilu yang memiliki integritas, profesionalitas, dan

^{1,2,3,4}Program Studi Magister Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim
 email: dwipurboyuwono@gmail.com

akuntabilitas. Dalam hal ini, tahun politik menjadi sebuah tahun yang seringkali ditunggu oleh sebagian masyarakat karena setiap mendekati tahun politik di setiap gelarannya pasti akan banyak sebuah fenomena atau peristiwa politik yang unik terjadi seperti di tahun-tahun sebelumnya salah satunya adalah komentar-komentar yang beragam dari masyarakat terutama dalam hal penentuan Presiden dan Wakil Presiden 2024. Dalam memasuki tahun-tahun, Dalam memasuki tahun-tahun politik saat ini akan banyak bertebaran informasi dan komentar dari masyarakat terhadap pelaksanaan pemilu, komentar atau pendapat yang disampaikan akan sangat beragam dimulai dari dukungan terhadap pelaksanaan pemilu, penggiringan opini publik, ujaran kebencian dan komentar-komentar lainnya (Manullang dan Prianto, 2023:12). Kemajuan teknologi saat ini mengakibatkan penyampaian pendapat dapat dengan mudah dipublikasikan melalui media sosial, salah satunya adalah melalui media twitter, twitter menjadi salah satu media sosial yang paling sering digunakan masyarakat dalam mengemukakan pendapatnya karena dianggap bebas (Buntoro, 2017:57).

Kemajuan teknologi saat ini mengakibatkan penyampaian pendapat dapat dengan mudah dipublikasikan melalui media sosial, salah satunya adalah melalui media twitter, twitter menjadi salah satu media sosial yang paling sering digunakan masyarakat dalam mengemukakan pendapatnya karena dianggap bebas

Hal ini sejalan dengan penyampaian suatu berita yang bersifat opini publik juga dijelaskan dalam ayat Al-Quran pada surat An-Nissa:83, sebagai berikut;

وَإِذَا جَاءَهُمْ أَمْرٌ مِّنَ الْأَمْنِ أَوْ الْخَوْفِ أَدَّعَوْا بِهٖ وَلَوْ رَدُّوهُ إِلَى الرَّسُولِ وَإِلَى أُولَى الْأَمْرِ مِنْهُمْ لَعَلَّمَهُ
وَلَوْ لَا فَضْلَ اللَّهِ عَلَيْكُمْ وَرَحْمَتَهُ لَاتَّبَعْتُمُ الشَّيْطَانَ إِلَّا قَلِيلًا الَّذِينَ يَسْتَنْبِطُونَهُ مِنْهُمْ

Artinya: Dan apabila datang kepada mereka suatu berita tentang keamanan ataupun ketakutan, mereka lalu menyiarkannya. Dan kalau mereka menyerahkannya kepada Rasul dan Ulil Amri di antara mereka, tentulah orang-orang yang ingin mengetahui kebenarannya (akan dapat) mengetahuinya dari mereka (Rasul dan Ulil Amri). Kalau tidaklah karena karunia dan rahmat Allah kepada kamu, tentulah kamu mengikut syaitan, kecuali sebahagian kecil saja (di antaramu).

Media twitter menjadi media yang sangat bebas dan luas dalam menyampaikan informasi sehingga dalam memasuki tahun politik ini akan ada banyak asumsi dari masyarakat terhadap pilpres 2024, komentar-komentar yang diberikan sangatlah beragam sehingga menimbulkan fenomena politik, dikarenakan komentar yang diberikan setiap individu unik dan beragam maka diperlukan sebuah analisa publik terkait pendapat pelaksanaan pilpres 2024, dengan melakukan analisis opini masyarakat yang menunjukkan pada sentimen masyarakat di media sosial twitter terhadap pilpres 2024 dapat mengetahui respon masyarakat Indonesia terhadap gelaran tersebut seperti apa, selain itu dapat dijadikan acuan untuk terhindar dari berita hoax, penggiran opini negatif, dan ujaran kebencian kepada satu pihak.

Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Pada gambar 1.2 terlihat bahwa banyak cuitan opini, serta sentimen dari masyarakat yang mengungkap pendapat dan pemikirannya. Oleh karena itu, diperlukan sebuah penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas opini dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari text mining, dalam hal ini adalah klasifikasi dokumen (Dewi dkk, 2023:12).

Berdasarkan latar belakang di atas, maka permasalahan yang dapat diidentifikasi dalam penelitian ini adalah

1. Berapa akurasi pemetaan opini masyarakat mengenai pilpres 2024 melalui X Twitter dengan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Self-Organizing Map?
 2. Apa saja keanggotaan hasil kluster dalam pemetaan opini masyarakat terhadap pilpres 2024 berdasarkan metode hasil dari Term Frequency-Inverse Document Frequency (TF-IDF) dan hasil pemetaan Self-Organizing Map?
- Sehingga akan mendapatkan tujuan penelitian untuk mengetahui berapa akurasi pemetaan

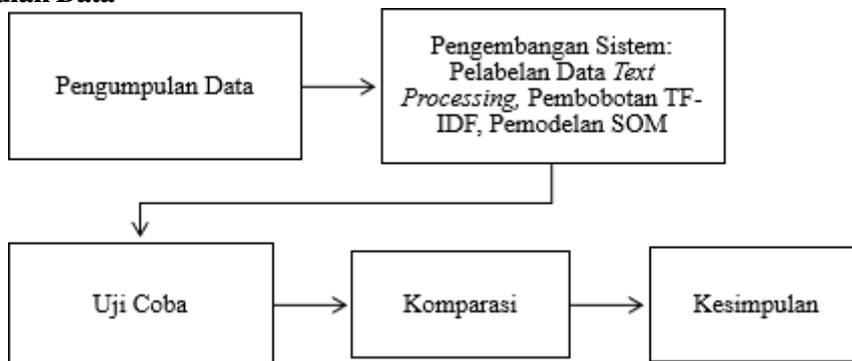
opini masyarakat mengenai pilpres 2024 melalui X Twitter dengan metode Term Frequency- Inverse Document Frequency (TF-IDF) dan Self-Organizing Map. Dan mengetahui keanggotaan hasil kluster dalam pemetaan opini masyarakat terhadap pilpres 2024 berdasarkan metode hasil dari metode Term Frequency-Inverse Document Frequency (TF-IDF) dan hasil pemetaan Self-Organizing Map.

METODE

1. Alur Penelitian

Penelitian diawali dengan melakukan pengumpulan data yang dilakukan dengan cara crawling data dari jejaring sosial Twitter untuk mendapatkan dataset berdasarkan kata kunci, pilpres2024, yang berupa cuitan dari pengguna atau media massa yang menyoal tentang sentimen dan tanggapan terhadap pilpres 2024. Dataset yang telah terkumpul akan dilakukan pelabelan terlebih dahulu untuk menentukan polaritas dari cuitan tersebut apakah positif, netral, atau negatif. Proses selanjutnya yang dilakukan adalah tahap pre-processing yang terdiri dari cleaning, case folding, tokenizing, normalization, filtering, dan stemming.

Pengumpulan Data



Menilik gambar desain sistem penelitian diatas merepresentasikan alur penelitian

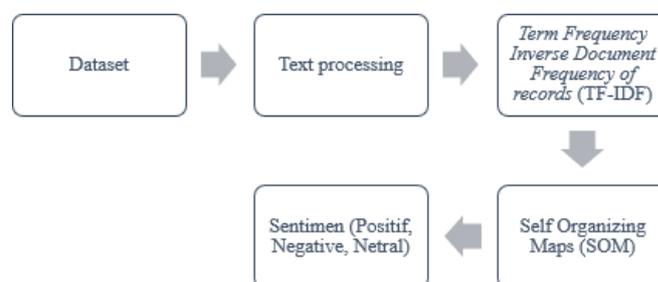
yang diawali dengan pengumpul-an data twitter, dataset Twitter yang berisi tweet-tweet terkait Pemilihan Presiden 2024 dikumpulkan menggunakan API Twitter atau alat pengumpulan data serupa. Data yang dikumpulkan mencakup teks tweet, metadata, tanggal posting, dan informasi lainnya yang relevan. Sebagai contoh

```

[
{
  "username": "user1",
  "tweet_text": "Mengapa harus memilih calon A? Karena dia memiliki pengalaman dan visi yang jelas untuk masa depan negara kita! #Pilpres2024",
  "timestamp": "2024-04-25 08:30:00",
  "retweets": 15,
  "favorites": 30,
  "hashtags": ["Pilpres2024", "CalonA"]
},
{
  "username": "user2",
  "tweet_text": "Saya tidak setuju dengan gagasan yang diajukan oleh calon B. Harus lebih teliti dalam memilih pemimpin! #Pilpres2024",
  "timestamp": "2024-04-25 09:15:00",
  "retweets": 8,
  "favorites": 20,
  "hashtags": ["Pilpres2024", "CalonB"]
}
]
  
```

Pengembangan Sistem

Pengembangan sistem dalam konteks pemrosesan teks melibatkan beberapa langkah penting, termasuk pelabelan data, pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), dan pemodelan menggunakan Self-Organizing Map (SOM).



Text Preprocessing

Proses selanjutnya adalah Text Preprocessing dengan berbagai tahapan yaitu.

a) Cleaning

Cleaning membersihkan data teks yang akan diolah. Data teks mentah yang masih banyak mengandung emoji, @username_pengguna akan dilakukan penghapusan. Untuk membersihkan data teks yang mengandung emoji dan @username_pengguna, dengan menggunakan beberapa teknik dalam proses cleaning. Berikut adalah rumus desain model dan contohnya terhadap pilpres 2024.

b) Case Folding

Case folding bertujuan membuat sebuah kalimat menjadi huruf kecil semua, hal ini agar mudah untuk memproses tahapan selajutnya. Dalam case folding, tujuannya adalah untuk membuat semua huruf dalam sebuah teks menjadi huruf kecil. Ini memudahkan proses selanjutnya dalam analisis teks karena menghilangkan perbedaan antara huruf besar dan huruf kecil.

c) Tokenizing

Tokenizing merupakan proses memecah kalimat menjadi kata-kata. Proses tokenizing biasanya mencakup penghapusan angka, tanda baca, simbol serta spasi yang tidak diperlukan. Dalam tokenizing, tujuannya adalah untuk memecah kalimat menjadi token-token atau kata-kata. Proses ini sering melibatkan penghapusan angka, tanda baca, simbol, dan spasi yang tidak diperlukan.

d) Normalization

Normalization atau Normalisasi merupakan tahapan yang dilakukan untuk mengubah kata yang tidak sesuai dengan aturan baku menjadi bentuk baku sesuai ejaan bahasa Indonesia. Dalam normalisasi teks, tujuannya adalah untuk mengubah kata-kata yang tidak sesuai dengan aturan baku menjadi bentuk yang baku sesuai dengan ejaan bahasa Indonesia.

e) Filtering

Filtering bertujuan untuk memisahkan antara kata hubung dan kata benda atau yang biasa disebut dengan penghapusan stopwords. Dalam filtering atau penghapusan stopwords, tujuannya adalah untuk memisahkan kata-kata yang tidak memberikan kontribusi signifikan dalam pemahaman teks, seperti kata hubung dan kata bantu. Menggunakan daftar kata-kata stopwords yang telah ditentukan sebelumnya dalam proses ini.

f) Stemming

Stemming merupakan tahapan yang dilakukan untuk mengubah kata yang memiliki imbuhan menjadi kata dasar. Biasanya kata kerja yang memiliki imbuhan seperti meng-ambil akan diubah menjadi kata dasarnya yaitu ambil. Dalam stemming, tujuannya adalah untuk mengubah kata-kata menjadi bentuk dasarnya atau kata dasar. Ini membantu dalam mengurangi dimensi data dan meningkatkan konsistensi, terutama dengan menghapus imbuhan atau akhiran kata.

HASIL DAN PEMBAHASAN

Metode Tf-Idf (Term Frequency-Inverse Document Frequency)

Simpulan yang dapat diambil terkait penerapan metode TF-IDF pada ringkasan otomatis dokumen Bahasa Indonesia untuk jenis fiksi dan non-fiksi berbasis bobot kata terhadap dokumen uji dapat dibagi dalam tiga bagian, berdasarkan compression rate, kepuasan

pembaca, dan durasi komputasi. Simpulan evaluasi secara detail dapat dilihat pada table 2.

Tabel 1. Simpulan Uji Coba

NO	Keterangan	Fiksi	Non-Fiksi	Keseluruhan	Satuan
1	Kalimat	44,5	19,25	31,875	kalimat
2	Kata	474,25	300,5	387,375	Kata
3	Durasi	79,5	57	68,25	detik

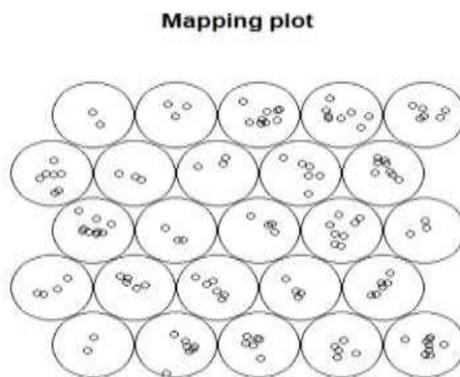
Metode Som (Self-Organizing Maps)

Untuk dapat menjalankan algoritma SOM pada RS tudio dibutuhkan library dengan nama “kahonen”. Selanjutnya menggunakan dataset yang sama maka diambil variable yang sama. yang penulis gunakan untuk melakukan clustering menggunakan algoritma SOM dan dilakukan proses standarisasi data. Sebelum proses pengelompokan menggunakan SOM dilakukan makan perlu untuk mengetahui berapa cluster optimal. Untuk mendapatkan jumlah cluster optimal maka dilakukan sebuah validasi internal berdasarkan nilai Dunn, Silhouette dan nilai Connectivity. Penentuan akan melihat berapa mana nilai Dunn yang paling mendekati nilai 1, lalu Silhouette tertinggi dan nilai Connectivity terendah. Dari hasil proses validasi yang dilakukan di Rstudio didapati hasil seperti pada tabel berikut.

Tabel 2 Nilai Connectivity, Dunn, dan Silhouette

Validation measure	Cluster Size								
	2	3	4	5	6	7	8	9	10
Connectivity	33.3028	32.7821	60.7877	65.9262	60.9825	81.3774	75.6893	100.8655	120.1306
Silhouette	0.0984	0.1637	0.0999	0.1270	0.1270	0.1270	0.1622	0.2062	0.1145
Dunn	0.2825	0.2287	0.2102	0.1990	0.2221	0.2074	0.2620	0.2350	0.1873

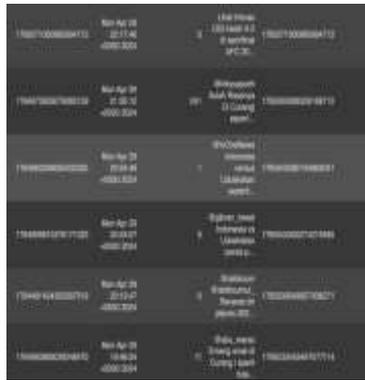
untuk mempermudah dalam membaca data tersebut, maka disajikan dalam bentuk plot awal seperti berikut.



Gambar 1 Plot tampilan awal dari data grid 5x5

Komparasi Performa Algoritma

Penelitian ini menggunakan dataset yang berasal dari TF-IDF beserta SOM dan dari 170 atribut dalam jangka waktu 3 bulan. Namun data-data tersebut masih terdapat data yang redundant, noisy, dan masih banyak missing value. Untuk keterangan tentang atribut dan deskripsi lengkap dapat dilihat pada penelitian (Strack, dkk., 2014). Tahap preprocessing perlu dilakukan karena data yang akan diadopsi dari UCI TF-IDF tersebut sebagai acuan perhitungan performa memiliki nilai atau data yang bersifat redundant, noisy, multiple (ganda), dan tidak komplit. Kemudian tahap selanjutnya adalah analisis missing values. Pada dataset yang tersedia terdapat empat jumlah atribut yang record atau datanya mengandung missing values.



Gambar 2. komparasi performa

SIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan bahwa Pertama, algoritme TF-IDF dapat digunakan untuk model klasifikasi sentimen cuitan pengguna Twitter dengan 70% data training dan 30% data testing didapatkan rata-rata akurasi 81%. Kedua, hasil pengujian masing-masing model dengan menggunakan confusion matrix didapatkan rata-rata accuracy sebesar 81%, precision 80,3%, recall 81%, dan f-measure 80%. Ketiga, berdasarkan perhitungan nilai preference value didapatkan hasil Prabowo Subianto 0,31 atau 31%, Anies Baswedan 0,29 atau 29%, Ganjar Pranowo 0,23 atau 23%. Sehingga dapat diketahui Prabowo Subianto memiliki elektabilitas paling baik berdasarkan cuitan Twitter bulan Juli 2022.

DAFTAR PUSTAKA

- Al-Quran Kementerian Agama RI. 2015. "Alquran dan Terjemahannya", Jakarta : Lajnah. Pentashihan Mushaf Al-Quran.
- Buntoro, Ghulam Asrofi. 2017. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter." *INTEGER: Journal of Information Technology* 2(1).
- Dewi, Anggi Riantika, Sri Diana, Moch Alvi Fakhrezi, Nana Awang, Helmi Ma'arif, dan Dedi Dwi Saputra. 2023. "Sentimen Analisis Terhadap Puan Maharani Sebagai Kandidat Calon Presiden 2024 Berdasarkan Opini Twitter Menggunakan Metode Naive Bayes Dan Adaboost." *JSiI (Jurnal Sistem Informasi)* 10(1):75–80.
- Manullang, Oktaviani, dan Cahyo Prianto. 2023. "Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden: Systematic Literature Review." *Jurnal Informatika dan Teknologi Komputer (J-ICOM)* 4(2):104–13.
- Poole, Jeanne E., Tristram D. Bahnson, Kristi H. Monahan, George Johnson, Hoss Rostami, Adam P. Silverstein, Hussein R. Al-Khalidi, Yves Rosenberg, Daniel B. Mark, dan Kerry L. Lee. 2020. "Recurrence of atrial fibrillation after catheter ablation or antiarrhythmic drug therapy in the CABANA trial." *Journal of the American College of Cardiology* 75(25):3105–18.
- Rahmah, Devi Amelia. 2022. "Analisis kluster berdasarkan indikator kesejahteraan rakyat menggunakan metode Self Organizing Maps (SOM)."
- Strack, B., DeShazo, J.P., Gennings, C., Olmo, J.L., Ventura, S., Cios, K.J., Clore, J.N. 2014. "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records". *BioMed research international Hindawi*, pp.1-11.