

Analisis Sentimen Pengguna Twitter dalam Pemilihan Presiden (PILPRES) 2024 dengan Menggunakan Algoritma K-Means

Abdusy Syakur Amin^{1*}, Dede Kurniadi², Zein nurzaman³, Rifa sri nurfadillah⁴, Sarah Khoerunisa⁵, Nisrina Khaerunisa⁶, Rafi Nurkholiq Ajiz⁷, Tegar Hanafi Jembar⁸, Ridwan Nur Faisal⁹

¹Universitas Pasundan, Indonesia

^{2,3,4,5,6,7,8,9}Institut Teknologi Garut, Indonesia

***email:** syakur.amin@unpas.ac.id

Info Artikel

Dikirim: 12 Januari 2024

Diterima: 31 Mei 2024

Diterbitkan: 31 Mei 2024

Kata kunci:

Analisis Sentimen;

K-Means;

Pilpres 2024;

Politik;

Twitter.

ABSTRAK

Salah satu bentuk penegakan demokrasi yang dilakukan oleh Negara Kesatuan Republik Indonesia yaitu melalui penyelenggaraan pemilihan presiden atau sering dikenal juga dengan istilah pilpres. yang diadakan setiap lima tahun sekali untuk memilih Presiden selanjutnya. Kemudian disamping itu, dalam era digital ini masyarakat semakin aktif menggunakan media sosial untuk menyampaikan pandangan, pendapat, dan sentimen mereka terkait dengan pemilihan presiden. Menjelang pilpres 2024 banyak kalangan seperti partai politik, tim sukses, buzzer, dan para pendukung memanfaatkan media sosial sebagai media berkampanye untuk meningkatkan popularitas dan elektabilitas bakal calonnya. Salah satu media sosial yang banyak digunakan dalam media promosi partai politik adalah twitter. Yang digunakan oleh orang-orang untuk mengirimkan berbagai komentar yang bisa bersifat positif atau negatif mengenai pemilihan tersebut. Terkadang, orang juga mengeluarkan pendapat yang bersifat hoaks sebelum atau selama pemilihan berlangsung. Mengingat komentar-komentar di Twitter saat ini sulit untuk dikategorikan sebagai positif atau negatif, diperlukan analisis sentimen guna memahami sikap publik terhadap pemilihan presiden tersebut. Penelitian ini bertujuan untuk mengevaluasi dokumen teks dan menentukan apakah dokumen tersebut memiliki orientasi sentimen positif atau negatif. selain itu metode yang digunakan yaitu K-Means untuk melakukan clustering pada data. Hasil dari pembobotan ini berupa sentimen positif dan negatif. Data diambil dari Twitter mengenai pemilu presiden (pilpres) 2024 sebanyak 1015 data tweet.

1. PENDAHULUAN

Pemilihan calon presiden yang diselenggarakan setiap 5 tahun sekali merupakan bagian dari proses demokrasi khususnya di Negara Indonesia. Seorang politisi yang ingin mencalonkan diri sebagai presiden tentu akan memeriksa atau mempertimbangkan popularitasnya berdasarkan opini publik. Sebelum adanya teknologi, masyarakat mengungkapkan opininya melalui media cetak yang dimana tidak semua orang mempunyai kesempatan dan kemampuan dalam menulis dan menerbitkannya [1]. Hanya orang-orang tertentu yang dapat menyuarakan pendapatnya melalui media cetak dengan tujuan agar dapat dilihat oleh banyak orang, kini setiap orang dapat mengkomunikasikan opininya dan opininya dapat dilihat oleh banyak orang karena kemajuan teknologi komunikasi di Indonesia. Demokrasi sendiri sudah berjalan melalui beberapa media online. Dikarenakan perkembangan zaman yang sudah begitu pesat hingga merambah sampai perkembangan teknologi

komunikasi saat ini, mengubah kecenderungan masyarakat dalam mengekspresikan opininya pada sebuah media sosial [2].

Media sosial yang marak dikenal oleh masyarakat untuk mengajukan pendapatnya dikenal dengan Twitter. Twitter adalah sosial media yang didirikan oleh Jack Dorsey pada tahun 2006 [3]. Pada tahun 2019, terdapat 500 juta kicauan atau tweet oleh pengguna Twitter per hari, menurut siaran pers Twitter [4]. Lebih dari 500 juta Tweet telah digunakan untuk memposting dan berbagi informasi tentang pengguna, dan juga konten Tweet dapat mengekspresikan perasaan. Twitter merupakan situs web yang mempunyai layanan menyediakan kumpulan data opini dari orang-orang di seluruh dunia. Hasil dari penyaluran opini dan komentar, Tweet merupakan sumber informasi yang dapat digunakan untuk menganalisis opini publik terhadap institusi dan individu. Opini pada tweet ini dapat digunakan untuk melihat bagaimana sentimen berjalan [5]. Dalam konteks demokrasi, pemilihan presiden yang adil dan berbasis opini publik merupakan aspek penting. Namun, dalam skala yang sangat besar seperti Pilpres, sulit untuk secara manual menganalisis setiap tweet pengguna Twitter terkait dengan calon presiden. Jika dilakukan secara manual, analisis opini dan sentimen pengguna Twitter terkait dengan calon presiden pada Pilpres 2024 akan menjadi sulit dan memakan waktu. Kurangnya pemahaman yang mendalam tentang persepsi publik dapat berdampak negatif pada proses demokrasi dan mengurangi tingkat partisipasi publik dalam Pilpres. Oleh karena itu, perlu adanya pendekatan komputasional dan metode analisis sentimen yang efisien untuk mengidentifikasi pola opini dan sentimen pengguna Twitter terkait dengan calon presiden [6].

Pada penelitian sebelumnya yang dilakukan oleh [7] dengan judul “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019” berhasil menerapkan metode K-Means dan Naive Bayes Classifier dengan hasil akurasi rata-rata sebesar 93,35% dan error rate rata-rata sebesar 6,66%. Kemudian penelitian selanjutnya dilakukan oleh [8] dengan judul “Penggunaan Metode K-Means Pada Analisa dan Klasifikasi Capres 2019 Di Twitter” yang juga berhasil menerapkan metode K-Means yang mana didapatkan validitas menggunakan DBI sebesar 0,52. Lalu penelitian dengan judul “Analisis Sentimen Masyarakat pada Media Sosial Twitter Terhadap Partai Politik Peserta Pemilu 2019 Menggunakan Naïve Bayes Classifier” yang dilakukan oleh [9] didapatkan hasil sentimen positif sebesar 53,8% dan sentimen negatif 46,13% dengan rata-rata akurasi sebesar 78,03%. Penelitian selanjutnya dilakukan oleh [10] dengan judul “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes” didapatkan hasil evaluasi dari dataset pertama adalah 73,68 akurasi dan AUC 0,74 pada fold ke-7, dataset kedua adalah 71,43 untuk akurasi dan AUC 1,0 pada fold ke – 5, untuk dataset ketiga nilai akurasi yang didapat 60% dan AUC 0,92 pada fold ke-1, dan untuk dataset terakhir nilai akurasi yang didapatkan adalah 62,5% dan AUC 0,65 pada fold ke-3. dan yang terakhir penelitian yang dilakukan oleh [11] dengan judul “Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier” yang menghasilkan klasifikasi pada data uji dengan nilai akurasi sebesar 71%. Untuk nilai akurasi tiap sentimennya yaitu 71% untuk sentimen negatif dan 70% untuk sentiment positif.

Berdasarkan latar belakang dan rujukan penelitian diatas, pada penelitian ini penulis menganalisa sentimen masyarakat pada twitter dalam pemilihan presiden 2024 menurut pandangan masyarakat. Metode yang digunakan yaitu Naïve Bayes Classifier [12]. Dengan demikian penelitian ini bertujuan untuk mengidentifikasi pola opini dan sentimen yang muncul dalam tweet terkait dengan calon presiden dan diharapkan dapat memberikan pemahaman yang lebih baik tentang persepsi publik terhadap calon presiden dan membantu memperkaya diskusi demokratis dalam pemilihan presiden.

2. METODE PENELITIAN

K-means adalah salah satu algoritma clustering yang populer dalam analisis data. Tujuan utama dari algoritma K-means adalah untuk mengelompokkan data ke dalam kelompok atau cluster yang memiliki kemiripan antara satu data dengan yang lain [13]. Proses algoritma K-means dimulai dengan memilih jumlah kluster (K) yang diinginkan. Setiap cluster direpresentasikan oleh sebuah centroid, yaitu titik pusat dari kluster tersebut. Selanjutnya, langkah-langkah K-means dilakukan secara berulang hingga konvergen, yaitu ketika tidak ada lagi perubahan dalam posisi centroid [14].

K-Means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok [15]. Metode ini mempartisi data kedalam kelompok yang lain [16]. Adapun tujuan pengelompokan data ini adalah meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok. Dengan menggunakan teknik klustering dalam K-means, maka tahapan algoritma K-means adalah sebagai berikut:

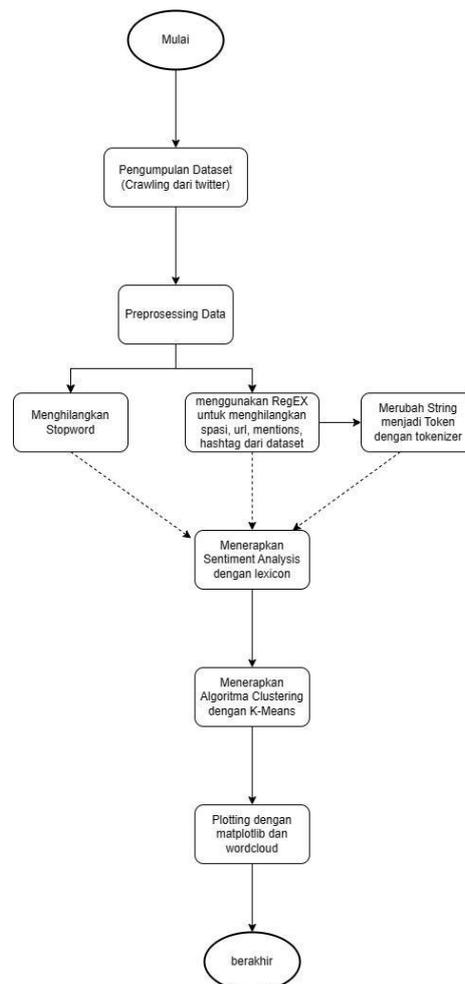
- Menentukan koordinat titik tengah setiap cluster
- Menentukan jarak setiap objek terhadap koordinat titik tengah
- Mengelompokkan objek-objek tersebut berdasarkan pada jarak minimumnya
- Tentukan pusat cluster baru
- Apakah ada selisih antar cluster lama dengan baru? jika masih ada kembali ke langkah a hingga d, jika tidak selesai.

Di dalam menentukan titik centroid kita menggunakan perhitungan jarak Euclidean distance, dengan rumus sebagai berikut :

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

x_i = koordinat x untuk fasilitas i
 y_i = koordinat y untuk fasilitas i
 d_{ij} = jarak antar fasilitas I dan j

Kemudian dalam penelitian ini dilakukan melalui beberapa tahapan yaitu sebagai berikut:



Gambar 1. Alur Proses Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data atau *Crawling Data*

Tweet atau kicauan publik yang dilakukan masyarakat di Twitter merupakan sumber data yang digunakan dalam penelitian ini, data tersebut diperoleh dengan menggunakan Rstudio dan Twitter API. Data yang terkumpul berjumlah sekitar 1015 data tweet, yang terdiri dari 7 atribut yaitu *created_at*, *id*, *reply_count*, *retweet_count*, *favourite_count*, *full_text*, *username*. Tabel *preprocessing* data dapat dilihat dalam tabel 1.

Tabel 1. *Preprocessing Data*

No	Created at	id	Full text	Reply count	Retweet count	Favourite count	username
0	Thu May 04 23:58:33 +0000 2023	1.654274e+18	Ganjar Ketar-ketir, Prabowo Diramal Bakal Mena...	0.0	14.0	18.0	Opsisicerdas
1	Thu May 04 23:57:54 +0000 2023	1.654274e+18	Pak Ganjar Pranowo bersilaturahmi ke kediaman ...	2.0	34.0	37.0	B1nt4g_K1r4r4
2	Thu May 04 23:59:26 +0000 2023	1.654275e+18	Pasca Deklarasi, Elektabilitas Ganjar Pranowo ...	0.0	0.0	0.0	adienugroho
...
...
1013	Thu May 04 15:59:47 +0000 2023	1.654154e+18	Bersatu, dukung upaya ini, mari berkontribusi ...	0.0	0.0	0.0	kolega_ganjar
1014	Thu May 04 16:00:39 +0000 2023	1.654154e+18	cuma Ganjar Pranowo yang mendapat dukungan dar...	0.0	0.0	0.0	andit1mo
1015	Thu May 04 15:59:33 +0000 2023	1.654154e+18	@S4N_W1B1 @jokowi @ganjarpranowo Ganjar peneru...	0.0	1.0	3.0	akupadi5

3.2 *Preprocessing Data*

Data yang didapatkan dari hasil pengumpulan data belum bisa langsung dikelompokkan karena data tersebut masih terdapat banyak simbol dan kata-kata yang tidak diperlukan atau dengan kata lain data tersebut masih berupa data mentah. Oleh karena itu, diperlukan *preprocessing* agar data lebih terstruktur dan bersih sehingga lebih mudah untuk dikelompokkan.

Tahapan *preprocessing* yang dilakukan adalah:

- 1) Menghilangkan Stopword: dilakukan dengan cara menghilangkan kata yang tidak diperlukan. Jika kata tersebut dibuang maka tidak akan mengubah atau menghilangkan informasi yang berada dalam kalimat tersebut. Seperti kata hubung yang, akan, di, pada dan lain-lain
- 2) Normalisasi: dilakukan dengan menggunakan RegEX untuk membersihkan fitur-fitur yang tidak diperlukan seperti Spasi, URL, Username, Hashtag dan lain-lain.
- 3) Tokenizing: dilakukan untuk memenggal kalimat menjadi beberapa bagian atau kata berdasarkan tanda bacanya seperti koma, titik, dan tanda baca lainnya.

Dataset yang sudah diberishkan dan dikelompokkan dapat dilihat pada tabel 2.

Tabel 2. *Dataset*

Created at	id	reply count	Retweet count	Favourite count	username	Clean text	<i>preprocessing</i>
Thu May 04 23:58:33 +0000 2023	1.654274e+18	0.0	14.0	18.0	Opsisicerdas	ganjar ketarketir prabowo diramal bakal menang...	[ganjar, ketarketir, prabowo, diramal, menang,...
Thu May 04 23:57:54 +0000 2023	1.654274e+18	2.0	34.0	37.0	B1nt4g_K1r4r4	pak ganjar pranowo bersilaturahmi ke kediaman ...	[ganjar, pranowo, bersilaturahmi, kediaman, kh..
Thu May 04 23:59:26 +0000 2023	1.654275e+18	0.0	0.0	0.0	adieenugroho	pasca deklarasi elektabilitas ganjar pranowo m...	[pasca, deklarasi, elektabilitas, ganjar, pran...
Thu May 04 23:58:43 +0000 2023	1.654274e+18	0.0	0.0	0.0	akakaka91413121	app meningkatnya kualitas pendidikan buat mas...	[app, meningkatnya, kualitas, pendidikan, masy...

3.3 Menerapkan *Sentiment Analysis* dengan Lexicon

Dalam analisis sentimen menggunakan lexicon positif dan negatif, polaritas teks dapat diklasifikasikan menjadi 'positive', 'negative', atau 'neutral'. Skor sentimen dihitung dengan mengakumulasi skor positif dan negatif dari kata-kata yang cocok dengan lexicon.

```
positive    611
neutral     237
Negative     89
Name: polarity, dtype: int64
```

Berdasarkan hasil analisis sentimen menggunakan lexicon positif dan negatif pada teks yang dianalisis, ditemukan bahwa terdapat dominasi sentimen positif dengan jumlah kemunculan sebanyak 611 teks. Selain itu, terdapat juga sebanyak 237 teks yang diklasifikasikan sebagai sentimen netral. Sementara itu, jumlah teks yang memiliki sentimen negatif adalah 89.

Hal ini menunjukkan adanya kecenderungan umum dari teks yang dianalisis untuk memiliki sentimen positif, diikuti oleh sentimen netral. Meskipun terdapat jumlah yang lebih kecil, tetap ada sejumlah teks yang mengungkapkan sentimen negatif.

Namun, perlu diingat bahwa analisis sentimen menggunakan lexicon memiliki batasan dalam mempertimbangkan konteks dan kemungkinan makna ganda atau ironi dalam teks. Oleh karena itu, hasil ini hanya memberikan gambaran umum tentang distribusi sentimen dalam teks yang dianalisis.

3.4 Melakukan Clustering dengan K-Means

Dalam konteks penelitian ini, clustering digunakan untuk mengelompokkan tweet-tweet yang berkaitan dengan Pilpres 2024 berdasarkan pola sentimen yang terkandung dalam teks tweet tersebut. Algoritma K-Means merupakan salah satu algoritma clustering yang umum digunakan dalam penelitian analisis sentimen. Proses clustering dengan K-Means dimulai dengan memilih jumlah kelompok atau kluster yang diinginkan. Dalam konteks ini, jumlah kluster dapat dipilih berdasarkan pertimbangan peneliti. Setelah itu, algoritma K-Means akan secara iteratif mengelompokkan tweet-tweet ke dalam kelompok-kelompok berdasarkan jarak euclidean antara fitur-fitur yang mewakili tweet tersebut.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter dalam konteks Pemilihan Presiden (Pilpres) 2024. Metode yang digunakan adalah algoritma K-Means untuk mengelompokkan tweet-tweet berdasarkan pola sentimen yang terkandung dalam teks tweet tersebut. Dengan menggunakan algoritma K-Means, penelitian ini berhasil mengidentifikasi pola sentimen dominan yang muncul dari pengguna Twitter terkait Pilpres 2024. Hasil analisis sentimen memberikan pemahaman yang lebih dalam mengenai pandangan, pendapat, dan perasaan pengguna terhadap calon presiden yang berpartisipasi dalam pemilihan tersebut. Informasi yang diperoleh dari analisis sentimen ini memiliki implikasi penting dalam bidang politik. Kandidat presiden dan tim kampanye dapat memanfaatkan hasil analisis sentimen untuk memahami persepsi publik terhadap platform Twitter dan menyesuaikan strategi kampanye mereka. Informasi ini dapat membantu dalam merancang pesan, menjawab isu-isu yang muncul, dan mengoptimalkan komunikasi dengan pemilih. Selain itu, penelitian ini memberikan wawasan tentang bagaimana pengguna Twitter merespons dan berpartisipasi dalam perdebatan politik yang berkaitan dengan Pilpres 2024. Informasi ini dapat digunakan untuk memahami tren dan dinamika politik di media sosial serta memperoleh pemahaman yang lebih luas tentang preferensi dan sikap publik. Secara keseluruhan, analisis sentimen pengguna Twitter menggunakan algoritma K-Means memberikan manfaat dalam hal pemahaman opini publik, perencanaan strategi kampanye, dan pengambilan keputusan politik selama Pilpres 2024.

REFERENSI

- [1] R. Vindua and A. U. Zailani, “Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python,” vol. 10, no. 2, pp. 479–487, 2024, doi: 10.30865/jurikom.v10i2.5945.
- [2] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM,” vol. 7, no. 1, pp. 11–18, 2024.
- [3] A. Agrani and B. Rikumahu, “Perbandingan Analisis Sentimen Terhadap Digital Payment ‘Go-Pay’ Dan ‘Ovo’ Di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes Dan Word Cloud Comparison of Sentiment Analysis Against Digital Payment ‘Go-Pay’ and ‘Ovo’ in Social Media Twitter Using N,” *Agustus*, vol. 7, no. 2, p. 2534, 2020.
- [4] N. Haqqizar and T. N. Larasyanti, “Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes,” *Prosiding TAU SNAR-TEK 2019 Seminar Nasional Rekayasa dan Teknologi*, vol. 10, no. 2, pp. 1–15, 2019.
- [5] H. Irsyad and M. R. Pribadi, “Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means,” *KURAWAL Jurnal Teknologi, Informasi dan Industri*, vol. 3, no. 2, pp. 164–172, 2020.
- [6] A. Ramdhani, B. K. Muzadid, D. T. Alamanda, and F. F. Roji, “Persepsi Netizen Terhadap Kebijakan Pemerintah Pada Ojek Online Menggunakan Twitter Analysis,” *Seminar Nasional ...*, 2019.
- [7] I. Kurniawan and A. Susanto, “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019,” *Eksplora Informatika*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [8] A. Rifa’i, G. G. Setiaji, and V. Vydia, “Penggunaan Metode K-Means Fakultas Teknologi Informasi dan Komunikasi , Universitas Semarang,” *Pengembangan Rekaya dan Teknologi*, vol. 15, no. 1, pp. 43–47, 2019.
- [9] A. R. Adiati, A. Herdiani, and W. Astuti, “Analisis Sentimen Masyarakat Pada Media Sosial Twitter Terhadap Partai Politik Peserta Pemilihan Umum 2019 Menggunakan Naive Bayes Classifier,” *eProceedings ...*, vol. 6, no. 2, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9836/9697>
- [10] M. R. Fais Sya’ bani, U. Enri, and T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [11] M. T. Student *et al.*, “Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier,” *Front Neurosci*, vol. 14, no. 1, pp. 1–13, 2021.

- [12] R. Julianto, E. D. Bintari, and Indrianti, “Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naïve Bayesian Classification,” *Journal of Big Data Analytic and Artificial Intelligence*, vol. 3, no. 1, 2017.
- [13] T. I. Saputra and R. Arianty, “Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat,” *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 191–198, 2019, doi: 10.35760/ik.2019.v24i3.2361.
- [14] S. J. Patandianan, G. M. Putra, and H. S. Pakpahan, “Implementasi Metode K-Means untuk Pengelompokan Rekomendasi Tugas Akhir,” vol. 16, no. 1, 2021.
- [15] T. I. Saputra and R. Arianty, “Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat,” *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 191–198, 2019, doi: 10.35760/ik.2019.v24i3.2361.
- [16] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *Sains Komupter & Informatika*, vol. 5, no. November 2019, pp. 697–711, 2021.