

# **Klasifikasi Sentimen Pada Buzzer Politik Jelang Pemilu 2024 menggunakan Metode Lexicon-based**

**YOULLIA INDRAWATY NURHASANAH, MUHAMMAD THORIQ NAUFAL**

Fakultas Teknologi Industri, Informatika, Institut Teknologi Nasional Bandung  
Email : [youllia@itenas.ac.id](mailto:youllia@itenas.ac.id)

*Received* 17 Oktober 2024 | *Revised* 12 November 2024 | *Accepted* 31 Oktober 2024

## **ABSTRAK**

*Pemilu merupakan sebuah momen dimana masyarakat memiliki peran untuk berpartisipasi dalam pemerintahan. Media sosial, khususnya media sosial X menjadi salah satu yang diminati untuk menyebarkan informasi. Informasi yang tersebar dalam jumlah banyak sangat sulit untuk disaring dengan baik oleh masyarakat. Mempengaruhi opini publik dengan menggunakan kata-kata kasar, memanipulasi informasi, dan membuat konten negatif merupakan peran buzzer politik di dalam media sosial. Sehingga dibutuhkan alat untuk filterisasi sentimen publik yang beredar. Alat yang digunakan adalah analisis sentimen. Penelitian bertujuan untuk melakukan klasifikasi sentimen buzzer politik dalam media sosial X menggunakan model lexicon-based yang ditingkatkan kamusnya sesuai dengan studi kasus yang dilakukan (corpus-based). Dengan menggunakan 1031 data, didapatkan sentimen positif (bukan buzzer) sebesar 63.69%, sentimen negatif (buzzer politik) sebesar 31.94%, dan sentimen netral sebesar 4.37%. Sehingga disimpulkan bahwa model lexicon-based mampu menjelaskan sentimen X.*

**Kata Kunci:** Analisis sentimen, Lexicon-based, Buzzer, X, Pemilu

## **ABSTRACT**

*Election is a moment where people have a role to participate in the government. Social media, especially X social media, is one of the most popular ways to disseminate information. Information that is spread in large quantities is very difficult for the public to filter properly. Influencing public opinion by using harsh words, manipulating information, and creating negative content is the role of political buzzers in social media. So a tool is needed to filter public sentiment that is circulating. The tool used is sentiment analysis. The research aims to classify the sentiment of political buzzers in social media X using a lexicon-based model that is enhanced by the dictionary according to the case study conducted (corpus-based). By using 1031 data, positive sentiment (not buzzer) is 63.69%, negative sentiment (political buzzer) is 31.94%, and neutral sentiment is 4.37%. So it is concluded that the lexicon-based model is able to explain X sentiment.*

**Keywords:** Sentimen analyzed, Lexicon-based, Buzzer, X, Election

## 1. PENDAHULUAN

Pemilu atau pemilihan umum merupakan bagian utama dalam sistem demokrasi di seluruh dunia, termasuk Indonesia. Pemilu adalah momen ketika masyarakat dapat memilih pemimpin mereka dengan memberikan suara. Pemilu bukan hanya sebuah acara politik, tetapi juga sebagai hak warga negara untuk berpartisipasi dalam pemerintahan (**Purmanudin, 2023**). Oleh karena itu, dengan mengikuti pemilu, masyarakat mampu memilih pemimpin yang memperjuangkan hak mereka dan menentukan kebijakan negara.

Pada masa ini, kemajuan teknologi dapat mengubah sistem politik. Berkembangnya teknologi informasi dan komunikasi (TIK) menyebabkan orang-orang lebih banyak berbicara di internet, baik di pemerintahan, lembaga sosial politik, maupun masyarakat. Salah satu sarana komunikasi internet yang banyak digunakan adalah media sosial (**Saputra, 2021**). Peran media sosial menjadi sangat penting dalam menyebarkan informasi saat ini. Masyarakat sudah mampu mengakses berbagai informasi yang beredar di internet. Dalam isu politik, informasi yang beredar mampu dipengaruhi dan dimanipulasi oleh suatu kelompok atau individu yang disebut dengan *buzzer* (**Wulandari, Musqith, & Ayuningtyas, 2023**).

*Buzzer* adalah sekelompok orang yang memiliki "pekerjaan" untuk menyebarkan pendapat atau ide, terutama di media sosial atau lebih luas lagi media online (**Kurniawati, 2023**). Pada awalnya, *buzzer* biasa digunakan untuk memasarkan suatu produk. Namun, keterlibatan *buzzer* dalam dunia politik mulai mengubah persepsi masyarakat terhadap *buzzer*. Sehingga *buzzer* kini dianggap sebagai pihak yang dibayar untuk memproduksi suatu konten negatif di media sosial dan telah dicap negatif (**Camil, Attamimi, & Esti, 2017**). *Buzzer* juga dijelaskan sebagai akun yang membuat laporan negatif, mengumpulkan informasi rancu, hingga menggunakan kata-kata kasar untuk mempengaruhi opini masa dengan tujuan membenci lawan politik (**Umami & Al Qindy, 2023**). Sehingga, berdasarkan pengertian dari kedua penelitian tersebut, akun-akun di media sosial yang menghasilkan sentimen negatif atau menggunakan kata-kata kasar untuk mempengaruhi opini publik akan diinterpretasikan sebagai sentimen *buzzer* politik.

Media sosial X menjadi media sosial yang sering digunakan untuk berbagi opini, terlebih lagi X populer karena kemampuan untuk menyampaikan pendapat objektif tentang berbagai topik (**Vindua & Zailani, 2023**). Data dikumpulkan menggunakan teknik *crawling* dari media sosial X, teknik ini digunakan karena mengumpulkan banyak data seperti *username*, waktu pembuatan, dan sentimen. Selanjutnya dilakukan teknik penyiapan data (*pre-processing*), analisis sentimen, dan visualisasi data hasil.

Analisis sentimen merupakan ilmu pengetahuan yang digunakan untuk mengetahui suatu opini masyarakat yang biasanya melalui bentuk komentar atau tulisan. Tanggapan yang didapatkan nantinya akan diklasifikasikan ke dalam sentimen positif atau negatif (**Ramadhani & Suryono, 2024**). Dalam studi yang dilakukan, model *lexicon-based* digunakan untuk menganalisa sentimen publik. Model ini dimulai dengan membuat sebuah kamus kata *lexicon* dan membandingkannya dengan opini publik (**Azhar, 2017**). Tingkat akurasi model *lexicon-based* dipengaruhi oleh banyaknya kata di dalam kamus *lexicon*, semakin banyak kata maka akan semakin tinggi tingkat akurasinya. *Lexicon corpus-based* merupakan teknik untuk meningkatkan akurasi dengan menambahkan kata ke dalam kamus kata yang sudah ada, kata yang ditambahkan diambil dari data aktual yang digunakan sehingga kamus *lexicon* dapat memiliki kata informal, kata slang, dan sesuai dengan topik yang dibahas (**Darwich, Noah, Omar, & Osman, 2019**).

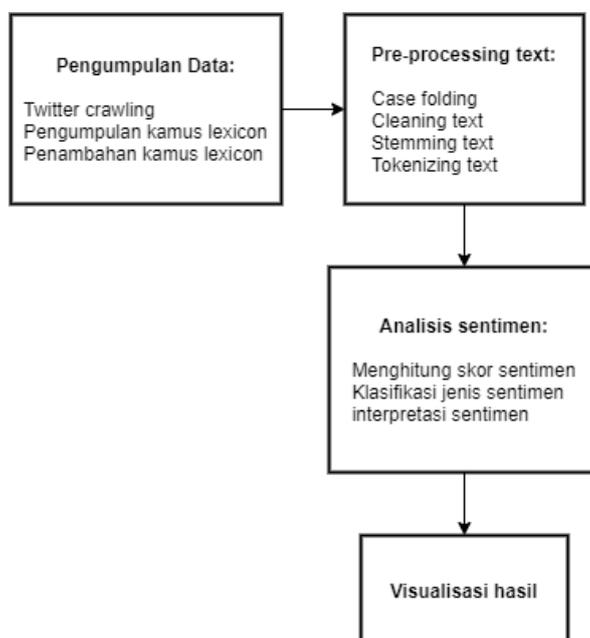
Metode *lexicon-based* umumnya masih diterapkan untuk pelabelan otomatis tanpa upaya signifikan untuk meningkatkan kualitas metode lexicon itu sendiri. Beberapa penelitian terbaru, seperti yang dilakukan oleh **(Saputri, Qur'ania, & Anggraeni, 2024)** dengan judul "Sentiment Analysis of Online Loans on Twitter Using Lexicon-Based Methods and Support Vector Machine (SVM)" dan **(Hendrawati, Ginantra, & Saiman, 2024)** dengan judul "Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vector Machine dan Lexicon-Based" menggabungkan metode *lexicon-based* dengan metode pembelajaran mesin (Machine Learning). Dalam penelitian-penelitian tersebut, metode lexicon digunakan untuk melakukan pelabelan data secara otomatis, yang kemudian digunakan dalam pemodelan berbasis pembelajaran menggunakan model SVM.

Namun, data aktual yang dihasilkan menggunakan metode *lexicon-based* masih menggunakan kamus lexicon yang lama. Akibatnya, metode lexicon yang bergantung pada kamus kata tidak dapat mengikuti perkembangan kata-kata baru atau spesifik untuk studi kasus tertentu. Oleh karena itu, sangat diperlukan peningkatan dan penyesuaian kamus lexicon agar lebih relevan dan sesuai dengan konteks studi kasus yang sedang diteliti.

Berdasarkan pemahaman di atas, dilakukan sebuah penelitian untuk melakukan klasifikasi buzzer politik dalam media sosial X menggunakan alat klasifikasi yaitu analisis sentimen dengan metode lexicon *corpus-based*. Hasilnya diharapkan dapat menentukan atau mengelompokkan buzzer berdasarkan opini-opini yang beredar di dalam media sosial X, sehingga masyarakat umum tidak mudah digiring ke dalam suatu opini dan dapat melakukan filterisasi terhadap suatu opini.

## 2. METODOLOGI PENELITIAN

Pada penelitian yang dilakukan, tahapan dibagi menjadi empat tahapan yang sistematis. Berikut beberapa tahapan yang dijelaskan dalam Gambar 1.



**Gambar 1. Tahapan Penelitian**

Tahap pertama adalah pengumpulan data yang dilakukan melalui *crawling* dari X, mengumpulkan kamus lexicon, dan menambahkan kamus lexicon sesuai dengan data aktual.

Selanjutnya, data dipra-proses sebagai teknik penyiapan data, antara lain *case folding*, *cleaning text*, *stemming text*, dan *tokenizing text*. Setelah itu, dilakukan analisis menggunakan metode *lexicon-based*. Metode *lexicon-based* akan menghasilkan nilai skor kalimat, klasifikasi jenis kalimat, dan interpretasi jenis kalimat. Tahap terakhir adalah melakukan visualisasi hasil data yang telah diperoleh.

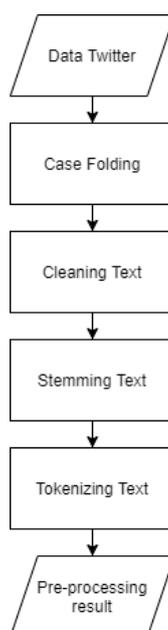
## 2.1. Pengumpulan Data

Pengumpulan data dilakukan dengan *crawling* data dari media sosial X. *Crawling* data dilakukan menggunakan teknik pengambilan sampel yang sesuai dengan topik politik, yaitu kata kunci terkait politik dalam bahasa Indonesia. Pengambilan data dilakukan pada tanggal 4 Februari 2024 sehingga sentimen terkait dengan isu pemilu sedang menjadi tren. Proses ini dilakukan menggunakan program karya Helmi Satria yang ditulis dalam bahasa pemrograman Python di Google Colab (**Satria, 2024**). Hasilnya diperoleh sebanyak 1031 sampel data primer dengan kata kunci "anies", "prabowo", "ganjar", dan "pemilu".

Selanjutnya dilakukan pengumpulan kamus *lexicon* positif dan kamus *lexicon* negatif. Peneliti menggunakan kamus *lexicon* bahasa Indonesia terbaru yang dibuat oleh (**Koto & Rahmaningtyas, 2017**), yang terdiri dari kamus *lexicon* positif sebanyak 3609 kata dan kamus *lexicon* negatif sebanyak 6609 kata.

Pada proses ini juga dilakukan penambahan kamus *lexicon* dengan membuat kamus *lexicon* yang sesuai dengan studi kasus, dalam konteks ini yaitu politik. Pembuatan kamus *lexicon* ini menggunakan metode *corpus-based*, di mana kamus *lexicon* dibuat berdasarkan data aktual yang ada sehingga kamus dapat mencakup kata informal, kata slang, dan sesuai dengan topiknya. (**Darwich, Noah, Omar, & Osman, 2019**).

## 2.2. Persiapan Data



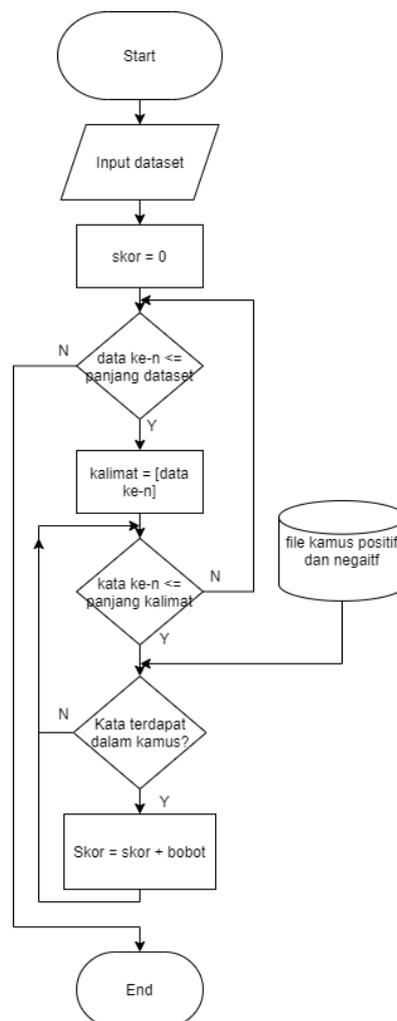
**Gambar 2. Pre-processing Data**

Persiapan data atau teknik *pre-processing* data merupakan tahapan untuk menyiapkan data mentah menjadi data yang siap olah. Ini diperlukan agar teks menjadi lebih sesuai untuk dianalisis atau proses selanjutnya. Pada penelitian ini, data hasil *crawling* akan disiapkan untuk

analisis sentimen. Terdapat beberapa tahap dalam pra-pemrosesan, seperti yang ditunjukkan pada Gambar 2.

### 2.3. Analisis Sentimen

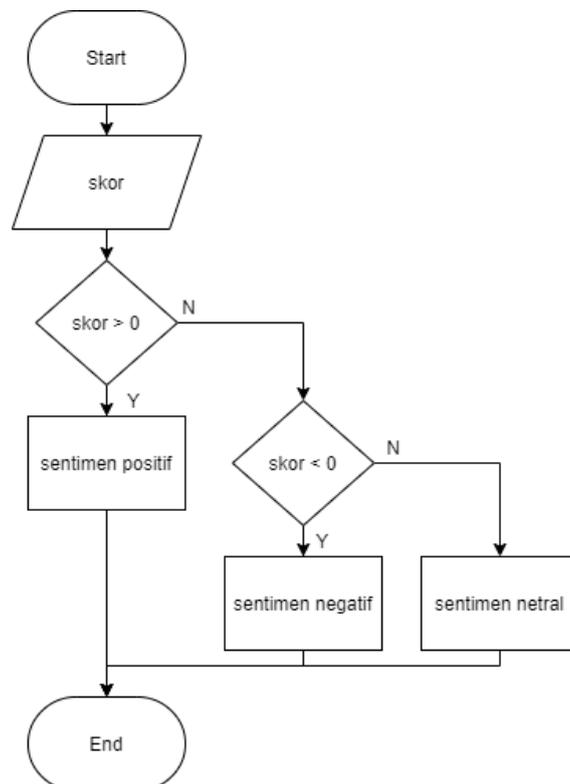
Analisis sentimen dapat dilakukan dengan pendekatan berbasis lexicon (lexicon-based) dan berbasis pembelajaran (*learning-based*) (Kusrini & Mashuri, 2019). Pada penelitian ini, analisis sentimen akan dilakukan menggunakan metode *lexicon-based*. Kamus yang digunakan merupakan hasil penelitian oleh (Koto & Rahmaningtyas, 2017), yang sudah ditingkatkan dengan menambahkan kamus sesuai studi kasus. Kamus terdiri dari dua kamus dasar yaitu kamus positif yang berisi kata dan bobot positif sebanyak 3819, serta kamus negatif yang berisi kata dan bobot negatif sebanyak 6782. Gambar 3 menjelaskan alur analisis sentimen menggunakan metode *lexicon-based*, di mana hasil pra-pemrosesan data yang berupa token-token akan dibandingkan dengan kamus kata yang telah ditingkatkan dengan menambahkan kata pada kamus lexicon. Skor akan dihitung dengan menjumlahkan bobot tiap kata dari satu kalimat. Proses ini akan dilakukan terus menerus hingga baris data terakhir, sehingga setiap baris akan memiliki bobot akhir.



**Gambar 3. Analisis Sentimen Menggunakan Lexicon-based**

Untuk mengetahui jenis kelas dari sebuah sentimen, dilakukan pembobotan lexicon, di mana proses ini memberikan nilai pada setiap sentimen (Farhan & Setiaji, 2023). Selanjutnya, hasil skor tiap data akan diklasifikasikan berdasarkan nilai skor tersebut. Jika nilai skor lebih

dari nol ( $> 0$ ) maka akan diklasifikasikan ke dalam kelas positif, jika nilai skor kurang dari nol ( $< 0$ ) maka akan diklasifikasikan ke dalam kelas negatif, dan jika skor sama dengan nol ( $= 0$ ) maka akan diklasifikasikan ke dalam kelas netral. Alur proses dari metode *lexicon-based* dapat dilihat pada Gambar 4.



**Gambar 4. Klasifikasi Sentimen**

Selanjutnya, berdasarkan pengertian *buzzer* dari (Wulandari, Musqith, & Ayuningtyas, 2023), yaitu kelompok yang aktif menggunakan media sosial untuk menyebarkan pesan politik hingga mempengaruhi opini publik, sentimen dengan pesan politik yang mampu mempengaruhi massa dapat diartikan sebagai *buzzer*. Selain itu, menurut (Umami & Al Qindy, 2023), *buzzer* adalah akun yang membuat laporan negatif, mengumpulkan informasi rancu, dan menggunakan kata-kata kasar untuk mempengaruhi opini massa dengan tujuan membenci lawan politik. Oleh karena itu, berdasarkan pengertian tersebut, akun yang membuat konten negatif atau sentimen negatif, membuat laporan negatif, dan berkata kasar untuk mempengaruhi opini di media sosial X akan diinterpretasikan sebagai *buzzer* politik.

#### 2.4. Visualisasi Hasil

Tahap terakhir dalam penelitian ini adalah visualisasi data yang dihasilkan. Hasil penelitian ini berupa dataset yang berisi interpretasi hasil analisis sentimen menggunakan metode *lexicon-based*. Data hasil akan divisualisasikan dalam bentuk diagram batang dan *word cloud*. Diagram batang akan menunjukkan persentase hasil interpretasi sentimen (Junianto, Arsi, Kusuma, & Saputra, 2024) sementara *word cloud* akan menggambarkan kata-kata yang paling sering muncul dalam data yang digunakan (Ananda & Suryono, 2024).





Kosa Kata	Bobot	Skor Akhir	Sentiment
'aja'	-2	4	
'ga'	-4	0	
'lindung'	3	3	
'malah'	-3	0	
'pegat'	-4	-4	
'noohhh'	-2	-6	
Keterangan			Negatif

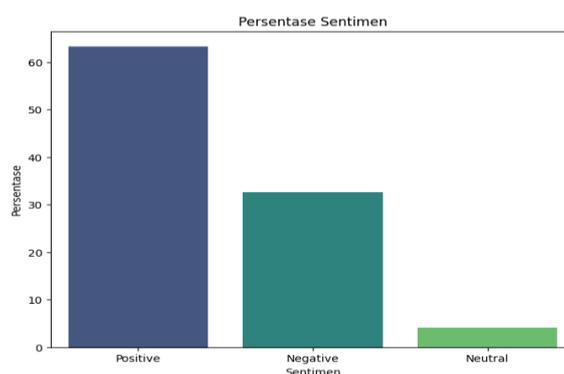
Setelah dilakukan klasifikasi kelas berdasarkan skor akhir kalimat, tahap berikutnya adalah interpretasi atau pemberian makna pada data. Dalam penelitian ini, yang menggunakan 1031 data, diperoleh hasil bahwa sebesar 63.30% sentimen adalah positif atau bukan buzzer, 32.62% sentimen adalah negatif atau buzzer, dan sisanya yaitu 4.08% sentimen adalah netral. Hasil analisis sentimen ini dapat dilihat pada Tabel 5.

**Tabel 5 Contoh Hasil Analisis Sentimen**

Preprocessing	Sentiment_score	Sentiment	Interpret
['idextratime', 'anies', 'tega', 'nanyain', 'soal', 'lindung', 'perempuan', 'ke', 'prabowo', 'lha', 'wong', 'prabowo', 'aja', 'ga', 'lindung', 'istri', 'malah', 'pegat', 'noohhh', 'titiék']	-6	Negatif	Buzzer
['pertanyaan', 'pak', 'anies', 'soal', 'lindung', 'kepada', 'perempuan', 'di', 'jawab', 'sendiri', 'oleh', 'pak', 'anies']	6	Positif	Bukan Buzzer
['rugi', 'sih', 'klo', 'pak', 'anies', 'ga', 'naik']	-5	Negatif	Buzzer
['wkwk', 'pa', 'anies', 'sblm', 'nanya', 'ke', 'pa', 'prabowo', 'senyam', 'senyum', 'rem', 'bgt']	0	Netral	Netral
['an**', 'finally', 'pak', 'anies', 'ngomongnya', 'lewat', 'waktu', 'walau', 'cma', 'terima', 'kasih', 'doang', 'wkakakakakaka', 'https', 't', 'co', 'vuafccpqu3']	-8	Negatif	Buzzer

### 3.4. Visualisai Hasil Data

Proses terakhir adalah visualisasi data untuk mempermudah pemahaman hasil analisis. Data akan divisualisasikan dalam bentuk diagram batang, seperti yang ditunjukkan pada Gambar 5, yang menggambarkan persentase hasil analisis. Selain itu, data juga akan divisualisasikan dalam bentuk *word cloud*, seperti yang ditunjukkan pada Gambar 6, yang menampilkan kata-kata yang paling sering muncul dari data yang digunakan.



**Gambar 5. Diagram Batang Hasil Analisis**





- Azhar, Y. (2017). Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Jurnal Nasional Pendidikan Teknik Informatika*, 238.
- Budiharto, W., & J, M. (2018). Prediction And Analysis Of Indonesia Presidential Election From Twitter Using Sentiment Analysis. *Journal Of Big Data*, 4.
- Camil, R., Attamimi, N. H., & Esti, K. (2017). *Di Balik Fenomena Buzzer: Memahami Lanskap Industri Dan Pengaruh Buzzer Di Indonesia*. Centre For Innovation Policy And Governance.
- Darwich, M., Noah, S. M., Omar, N., & Osman, N. A. (2019). Corpus-Based Techniques For Sentiment Lexicon Generation: A Review. *Journal Of Digital Information Management*, 296.
- Farhan, N. M., & Setiaji, B. (2023). Komparasi Metode Naive Bayes Dan SVM Pada Sentimen Twittermengenai Persoalanperppucipta Kerja. *Indonesian Journal Of Computer Science*, 2719.
- Hendrawati, T., Ginantra, N. W., & Saiman, C. M. (2024). Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine Dan Lexicon Based. *Tematik: Jurnal Teknologi Informasi Komunikasi*, 56-64.
- Junianto, H., Arsi, P., Kusuma, B. A., & Saputra, D. I. (2024). Evaluasi Aplikasi Raileo Melalui Analisis Sentimen Ulasan Playstore Dengan Metode Naive Bayes. *Science And Information Technology Journal*, 27-40.
- Koto, F., & Rahmaningtyas, R. Y. (2017). Inset Lexicon: Evaluation Of A Word List For. *International Conference On Asian Language Processing*, 391-393.
- Kurniawati, R. (2023). Buzzer Sebagai Alat Politik Ditinjau Dari Perspektif Penegakan Hukum Di Indonesia. *Justicia Sains: Jurnal Ilmu Hukum*, 260-275.
- Kusrini, & Mashuri, M. (2019). Sentiment Analysis In Twitter Using Lexicon Based And Polarity Multiplication. *International Conference On Applied Innovations In IT*, 365-368.
- Prasetya, Y. N., Winarso, D., & Syahril. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19. *JURNAL FASILKOM*, 97-103.
- Purmanudin, H. (2023, Oktober 7). *Pemilu: Pilar Demokrasi Dan Partisipasi Warga Negara*. Retrieved From Kompasiana: <https://www.kompasiana.com/hotim16665/6521379b110fce70ec734d73/Pemilu-Pilar-Demokrasi-Dan-Partisipasi-Warga-Negara>
- Ramadhani, B., & Suryono, R. R. (2024). Komparasi Algoritma Naïve Bayes Dan Logistic Regression Untuk Analisis Sentimen Metaverse. *Jurnal Media Informatika Budidarma*, 714.

- Saputra, E. (2021, Januari 6). *Peranan Media Sosial Di Dalam Bidang Politik*. Retrieved From Kumparan: <https://kumparan.com/erik-saputra-1609901036015096387/peranan-media-sosial-di-dalam-bidang-politik-1uvmyhjvhac>
- Saputri, C. S., Qur'ania, A., & Anggraeni, I. (2024). Sentiment Analysis Of Online Loans On Twitter Using Lexicon Based Methods And Support Vector Machine (SVM). *KOMPUTASI: Jurnal Ilmiah Ilmu Komputer Dan Matematika*, 37-43.
- Satria, H. (2024, Maret 30). *Crawl Data Twitter > 2000 Tweets - 30 Maret 2024*. Retrieved From Google Colab: [https://colab.research.google.com/drive/1rymmmol4e-3xnix\\_filfioqnytub-4m#scrollto=4UIL1x21P9rQ](https://colab.research.google.com/drive/1rymmmol4e-3xnix_filfioqnytub-4m#scrollto=4UIL1x21P9rQ)
- Umami, A. M., & Al Qindy, F. H. (2023). The Use of Buzzers by Political Parties Which Result in Black Campaign Practices in Indonesia. *Jurnal Ilmu Sosial dan Pendidikan (JISIP)*, 3365-3373.
- Vindua, R., & Zailani, A. U. (2023). Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter . *JURIKOM*, 479.
- Wulandari, C. D., Musqith, M. A., & Ayuningtyas, F. (2023). Fenomena Buzzer Di Media Sosial Jelang Pemilu 2024. *AVANT GARDE: Jurnal Ilmu Komunikasi*, 135-136.