



# Identifikasi Topik Dominan pada Hastag Konten Anak di Media Sosial X Menggunakan Word Cloud

Nisrina Akbar Rizky Putri<sup>1\*</sup>  
1 Universitas Muhammadiyah Klaten, Indonesia  
\*nisrinaakbar@umkla.ac.id

## INFORMASI ARTIKEL

### Sejarah Artikel:

Diterima Redaksi: 01 Mei 2025

Revisi Akhir: 26 Mei 2025

Diterbitkan Online: 26 Juni 2025

## KATA KUNCI

Word Cloud; Crawling Data; X; Analisis Topik

## KORESPONDENSI

Telepon: +6285726541777

E-mail: nisrinaakbar@umkla.ac.id

## ABSTRACT

Penelitian ini bertujuan untuk mengidentifikasi topik-topik dominan dalam konten bertagar anak (Konten Anak) di media sosial X. Data yang digunakan berupa 1.967 cuitan yang diperoleh melalui proses crawling dengan menggunakan kata kunci dan tagar yang relevan. Analisis dilakukan dengan pendekatan eksploratif menggunakan visualisasi Word Cloud untuk menampilkan kata-kata yang paling sering muncul dalam cuitan tersebut. Proses analisis meliputi langkah-langkah pra-pemrosesan seperti *transformation*, *tokenization*, dan *filtering* untuk meningkatkan akurasi visualisasi. Hasil penelitian menunjukkan bahwa kata-kata "anak-anak" dan "konten" mendominasi perbincangan, namun ditemukan pula kata-kata negatif seperti "fantasi", "sedarah", dan "pornografi" yang mengindikasikan adanya kemungkinan pelecehan atau eksploitasi konten anak di media sosial.

## 1. PENDAHULUAN

Dalam beberapa waktu terakhir, perkembangan tidak hanya dalam hal teknologi saja, akan tetapi penggunaan media sosial juga mengalami peningkatan yang signifikan. Pengguna media sosial merupakan sebuah fitur yang sangat diminati oleh masyarakat. Portal seperti X (sebelumnya Twitter), TikTok, Instagram dan Youtube sekarang menjadi integral bagian dari aktivitas sehari-hari baik dari segi hiburan, komunikasi atau pembelajaran. Tidak hanya dari kalangan orang dewasa saja, akan tetapi saat ini anak-anak dan remaja cukup tertarik pada media sosial yang dapat menampilkan video streaming secara langsung[1].

Kemudahan dalam mengakses internet dalam perangkat ini semakin mendorong terutama pada anak-anak untuk aktif berpartisipasi dalam dunia virtual sejak usia dini. Anak-anak saat ini sudah lebih pintar dalam menggunakan *smartphone* dan dapat mengakses media sosial dengan mudahnya. Dari sisi lainnya terdapat perkembangan bahasa dari perspektif linguistik, anak-anak kini mendapatkan stimulus bahasa yang sering kali lebih banyak didapatkan dari dunia digital. Hal ini diperoleh disaat anak-anak menggunakan media sosial untuk menjadi penonton pasif dengan menirukan gaya bicara, ekspresi ataupun kata-kata

yang digunakan oleh beberapa konten kreator favorit mereka[2]. Konten media sosial yang diaksespun sangatlah beragam, mulai dari dokumentasi aktivitas sehari-hari, video lucu, dan edukasi. Tidak sedikit juga terdapat konten yang menjadikan anak-anak viral karena mencuri perhatian, baik dari tingkah lakunya yang lucu, ekspresinya ataupun ada beberapa situasi yang dinilai kurang baik dikarenakan konten viral karena sebuah kontroversi.

Hal ini tentu saja menjadi masalah utama pada saat ini dikarenakan orang tua yang harus tahu cara dalam mengelola waktu penggunaan *smartphone* karena jika tidak memiliki kontrol yang baik dalam menggunakan *smartphone* dapat berimbas pada anak-anak mereka[3]. Salah satunya media sosial TikTok memiliki istilah seperti pisau bermata dua, dikarenakan TikTok memiliki dampak positif yang digunakan untuk wadah berkreasi, berimajinasi atau memperluas relasi sosial. Sedangkan tidak sedikit juga TikTok memiliki konten yang bersifat negatif[4]. Kekhawatiran publik saat ini meningkat dikarenakan adanya beberapa konten anak yang justru menampilkan sisi eksploitasi terselubung, seperti komentar negatif atau penyalahgunaan citra anak untuk keuntungan finansial. Anak-anak tidak memahami bahwa dampak digital dalam jangka waktu panjang dengan bisa mengalami masalah kesehatan mental anak termasuk seperti gangguan kecemasan, stres dan kecanduan *smartphone*[5]. Algoritma media sosial mendorong konten-

konten viral juga berperan dalam penyebarannya. Situasi ini menunjukkan pentingnya perhatian terhadap keamanan dan kenyamanan anak dalam ruang dunia maya.

Saat ini belum ada undang-undang yang kuat dalam mengatur dan memantau konten anak di media sosial terutama pada media sosial global. Andai undang-undang anak berlaku di beberapa negara, penggunaan dan penguatan perundang-undangan di bidang digital seringkali menantang. Banyak diantaranya hanya memproses laporan secara manual, sehingga konten yang merusak ataupun menjatuhkan mental anak-anak sering kali hanya dilewatkan. Dengan meneliti topik populer yang terkait dengan konten anak di media sosial X, studi ini bertujuan mengetahui kata apa yang sering bermunculan dalam pembahasan konten anak. Diharapkan bahwa topik-topik yang sering muncul dalam wacana digital mengenai anak-anak akan ditemukan melalui pendekatan eksploratif yang menggunakan teknik visualisasi *word cloud*. Temuan analisis akan menawarkan ringkasan awal tentang bagaimana masyarakat umum memandang konten anak dan topik-topik yang kini menjadi perhatian masyarakat. Diharapkan bahwa hasil ini akan menjadi dasar untuk penelitian tambahan dari sudut pandang kebijakan media, literasi digital masyarakat, dan perlindungan anak.

## 2. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Josua, dkk[6] melakukan klasifikasi sentimen ulasan pada *e-commerce* Shopee berbasis *word cloud* dengan metode *naive bayes* dan *k-nearest neighbor*. Peneliti menggunakan 500 ulasan dari website *google play store* dan mengelola ulasan tersebut dengan menggunakan aplikasi *orange* dengan menggunakan metode *word cloud* untuk mengetahui topik-topik yang sering diulas oleh pelanggan. Hasil yang diperoleh adalah berupa kata sentimen positif yang paling sering diulas oleh pelanggan diantaranya kata: gratis, bagus, suka, murah, mudah, dan cepat. Sedangkan sentimen negatifnya yang diperoleh berupa kata: kecewa, jelek, mahal, bohong, ribet, dan diperbaiki.

Socia dan Budi [7] melakukan analisis sentimen terhadap program merdeka belajar dengan *text analysis word cloud* dan *word frequency*. Penelitian ini menggunakan data kuesioner yang berjumlah 149 data responden. Para responden diberikan pertanyaan seputar 4 hal dan kemudian datanya akan diproses untuk memunculkan kata dengan frekuensi paling banyak digunakan. Hasil yang didapatkan RQ 1, RQ 2, dan RQ 3 banyak kata yang sama muncul yaitu “belajar. Dan pada RQ 4 memiliki perbedaan dengan RQ lainnya, RQ 4 kata yang sering muncul adalah “dosen”.

Penelitian yang dilakukan Irvandi dkk[8] menggunakan ulasan pada *google maps review* untuk mengetahui asumsi wisatawan terhadap wisata halal di Pulau Lombok. Dengan melakukan *web scraping* dan mengelolanya menggunakan aplikasi *rapidminer*, ulasan tersebut diproses dengan menggunakan algoritma *naive bayes* serta visualisasi menggunakan *word cloud*. Untuk memunculkan beberapa kata yang sering diutarakan wisatawan, dilakukan terlebih dahulu *text processing* berupa *tokenize*, *filter token by length* *transform case*, *stopword* dan *stemming*. Dari hasil algoritma *naive bayes* didapatkan akurasi sebesar 74,75%

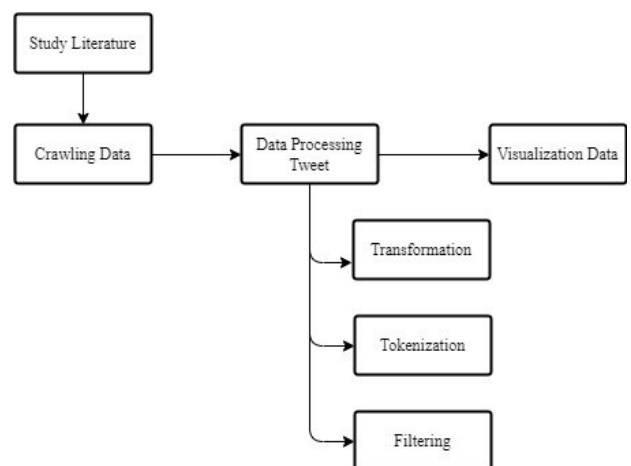
dan visualisasi *word cloud* berupa kata: indah, wisata, pantai, alam, gunung dan masjid.

Iqbal dan Ridwan [9] penelitian ini memfokuskan pada pemahaman kata kunci dalam strategi pemasaran secara online yang dinilai pada penggunaan *google custom search*, *spacy* dan *word cloud* untuk mencari tren dan topik utama dari hasil *googling*. Dikarenakan keterbatasan metode, peneliti merekomendasikan pengembangan analisis melalui perluasan cakupan, penggunaan visualisasi lanjutan serta kolaborasi ahli untuk meningkatkan efektivitas strategi konten dan SEO di era digital.

Zulham dkk [10] melakukan penelitian analisis sentimen masyarakat terhadap sepeda motor listrik di Twitter dengan menggunakan *orang data mining*. Dengan mendapatkan 2000 tweets yang didapatkan dari hasil *crawling*, tweet akan di proses untuk menghilangkan beberapa kata atau simbol yang tidak dibutuhkan atau terlalu sering muncul. Hasil visualisasi *word cloud* dengan kata yang paling sering muncul adalah kata: mobil listrik, motor dan sepeda.

## 3. METODOLOGI

Penelitian ini dilakukan dengan pendekatan eksploratif dari cuitan yang dapatkan dari hasil *crawling* melalui X. Pada Gambar 1 merupakan diagram alur penelitian. Dimulai dari melakukan studi literatur review terlebih dahulu untuk topik pembahasan yang memfokuskan pada konten anak dan *word cloud*. Kemudian dilakukan *crawling* data dengan melalui *google colab* dan *tools* yang digunakan berupa *tweet-harvest*. Selanjutnya dataset diproses dengan menggunakan aplikasi *orange* untuk dilakukan *transformation*, *tokenization*, dan *filtering*. Lalu hasil akhirnya yang didapatkan berupa visualisasi data *word cloud* untuk mengetahui kata apa saja yang menjadi pembahasan topik utama dalam dataset tweet tersebut.



Gambar 1. Bagan Alur Penelitian

### 3.1. Crawling Data

*Crawling data* bertujuan untuk mengumpulkan informasi dari berbagai sumber dengan bentuk file yang beragam, seperti *database* atau *application programming interface* (API).

Pengambilan data bertujuan untuk dapat melakukan analisis berupa kebutuhan dalam pengembangan dan penelitian[11]. Proses ini biasanya membutuhkan alat khusus yang membantu *crawling data* dan menjelajahi *website* yang dituju dan mengumpulkan data dalam jumlah besar[12]. Pada penelitian ini, untuk melakukan *crawling data* menggunakan *google colab* dan menggunakan *tweet-harvest* sebagai alatnya.

### 3.2. Orange Data Mining

*Orange* merupakan salah satu aplikasi *open-source* yang digunakan untuk *machine learning* atau penambahan data. *Orange* digunakan untuk menganalisis dan memvisualisasi data eksploratif dan menawarkan *platform* untuk permodelan prediktif, pemilihan eksperimen dan sistem rekomendasi[13]. Aplikasi ini memiliki salah satu widget yang bisa menampilkan kata-kata yang menjadi topik utama pembahasan pada suatu tweet yang disebut sebagai *word cloud*. Penelitian ini akan menggunakan widget tersebut untuk menunjukkan kata yang menjadi topik terbanyak yang dituliskan oleh masyarakat.

### 3.3. Processing Text (Tweet)

Proses ini merupakan pengolahan teks yang sudah didapatkan pada saat *crawling data* dikarenakan data tersebut masih *raw data*. Maka teks melalui proses transformasi dengan dibagi menjadi ukuran yang lebih kecil atau sama, *tokenization* yang memecah kalimat menjadi komponen yang lebih kecil, dan *filtering* yang merupakan tahap penghilangan kata atau tanda yang tidak digunakan[14].

### 3.4. Word Cloud

*Word Cloud* merupakan salah satu widget yang memvisualisasikan kata dengan menunjukkan daftar kata-kata yang digunakan dalam sebuah teks dalam bentuk abstraksi. Umumnya semakin banyak kata yang digunakan dalam sebuah teks, maka akan semakin besar bentuk representasi kata tersebut dalam gambar[15]. *Word cloud* dikenal juga sebagai *cloud text* atau *cloud tag* yang merupakan metode untuk menarik bagian teks paling relevan. *Word cloud* juga dapat membantu dalam membandingkan dan membedakan dua bagian teks untuk menemukan kata yang serupa.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Sumber Data Tweets

Pada tahapan ini berfokus pada pengumpulan data cuitan masyarakat dari media sosial X yang berkaitan dengan kata ‘Konten Anak’ dan ‘Konten Viral Anak’. Media sosial X menyediakan *API* dan juga *auth token* yang memungkinkan untuk *developer* atau peneliti melakukan akses terprogram pada media sosial tersebut, hanya saja saat ini jumlah data yang dapat diunduh dibatasi. Pengumpulan data *tweets* dilakukan dengan menggunakan *google colab* dan *tools tweet-harvest*. Pada Gambar 2 a dan b merupakan code untuk melakukan *crawling data* pada X dengan limit yang didapatkan berjumlah 1500 cuitan.

```
filename = 'konten_anak.csv'
search_keyword = 'konten anak since:2023-04-01 until:2025-06-01 lang:id'
limit = 1500

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}
```

(a)

```
filename = 'konten_viral_anak.csv'
search_keyword = 'konten viral anak since:2023-04-01
until:2025-06-01 lang:id'
limit = 1500

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}
```

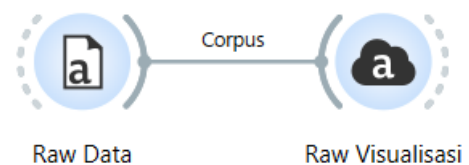
(b)

Gambar 2. Step Crawling Data

*Crawling data* dilakukan sebanyak 2 kali sehingga mendapatkan 2 dataset, dataset 1 mendapatkan cuitan sejumlah 1500 dan pada dataset 2 mendapatkan cuitan sejumlah 467. Masing-masing dataset memiliki 15 kolom yang beberapa nama kolomnya seperti *created\_at*, *favorite\_count*, dan *full\_text*. Pada penelitian ini, peneliti hanya akan menggunakan *full\_text* sebagai media yang digunakan agak bisa membentuk *word cloud* dan menunjukan kata yang sering digunakan oleh masyarakat.

### 4.2. Pre-Processing Data dan Visualization Word Cloud

Penelitian selanjutnya melakukan *merged* pada 2 dataset yang sudah didapatkan dan diambil kolom *full\_text* untuk mendapatkan teks yang dibutuhkan saat visualisasi data. Peneliti menggunakan aplikasi *Orange* untuk melakukan visualisasi dan pre-processing data. Pada Gambar 4 merupakan raw data yang langsung divisualisasikan. Pada Gambar 5 menghasilkan visualisasi dari banyak kata sambung, kata tidak baku dan simbol yang menjadi topik *word cloud*.



Gambar 3. Widget Raw Data





*Technology*, 2024, doi: 10.25077/Attribution-NonCommercial.

- [13] R. Aghnia, R. Wiguna, and A. I. Rifai, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining," *Journal of Information Systems and Informatics*, vol. 3, no. 1, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi>
- [14] H. N. Azizah, "VISUALISASI ANALISIS SENTIMEN SIBERBULLYING PADA POST INSTAGRAM MENGGUNAKAN ORANGE DATA MINING," *IJCSR: The Indonesian Journal of Computer Science Research*, vol. 1, no. 1, 2023, [Online]. Available: <https://subset.id/index.php/IJCSR>
- [15] T. M. Fahrudin *et al.*, "Analisis Speech-to-Text pada Video Mengandung Kata Kasar dan Ujaran Kebencian dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens dan Visualisasi Word Cloud," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 5, no. 2, pp. 190–202, 2022.