

## **Analisis Konten Hoaks di Indonesia: Tren Penyebaran dan Kontribusi Media Sosial**

**Joceline Schellenberg W**

Program Studi Informatika, Universitas Deztron Indonesia

Corresponding Author: [jocelineschellenberg@udi.ac.id](mailto:jocelineschellenberg@udi.ac.id)

---

### **Info Artikel**

**Submitted:** 05 Oktober 2025

**Revised :** 11 Oktober 2025

**Accepted:** 15 Desember 2025

**Published:** 27 Desember 2025

**Keywords:** Web Scraping; Hoax Trends; Social Media

**Kata Kunci:** Web Scraping; Tren Hoaks; Media Sosial

---

### **Abstract**

*The spread of false information (hoax and fake news) on social media represents a serious challenge to the integrity of public information in Indonesia. This research aims to analyze the trends, volume, and dissemination patterns of hoaks based on verification data collected via web scraping from local fact-checker websites during the period of January to August 2025. The results of the data analysis show that the volume of hoaks is significantly dominated by two categories: Factual Disinformation and Dubious Claims. Platform contribution analysis indicates that Facebook and Instagram are the main channels of dissemination for both categories. These findings provide important insights for stakeholders and social media platforms to formulate segmented and targeted mitigation strategies.*

---

### **Abstrak**

*Penyebaran informasi palsu (hoaks dan fake news) di media sosial merupakan tantangan serius terhadap integritas informasi publik di Indonesia. Penelitian ini bertujuan untuk menganalisis tren, volume, dan pola diseminasi hoaks berdasarkan data verifikasi yang dikumpulkan melalui web scraping dari situs fact-checker lokal selama periode Januari hingga Agustus 2025. Hasil analisis data menunjukkan bahwa volume hoaks didominasi secara signifikan oleh dua kategori: Disinformasi Faktual dan Klaim Meragukan. Analisis kontribusi platform menunjukkan bahwa Facebook dan Instagram menjadi saluran utama penyebaran di kedua kategori tersebut. Temuan ini memberikan wawasan penting bagi stakeholder dan platform media sosial untuk merumuskan strategi mitigasi yang tersegmentasi dan ditargetkan*



*This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).*

**Publisher: Lembaga Penerbit Penelitian Nusantara**

---

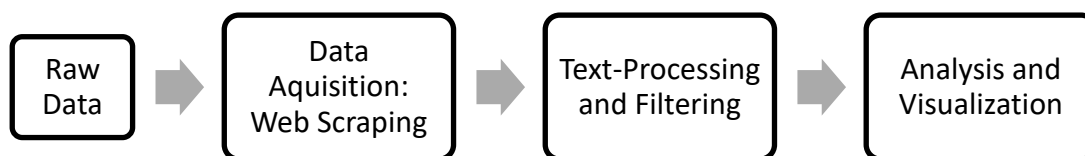
### **Pendahuluan**

Fenomena disinformasi atau fake news telah bertransformasi menjadi ancaman global, terutama di negara dengan tingkat penetrasi internet dan media sosial yang tinggi seperti Indonesia. Upaya untuk melawan banjir informasi palsu ini sangat bergantung pada data yang terverifikasi. Lembaga-lembaga fact-checker lokal berperan krusial dengan memverifikasi klaim publik dan mengategorikannya. Oleh karena itu, penelitian ini dilakukan untuk menggali pola tersembunyi dari data verifikasi tersebut, khususnya untuk memahami kapan (tren temporal) dan di mana (platform media sosial) hoaks paling aktif beredar di Indonesia. Pemanfaatan data fact-checker yang terbuka melalui teknik web scraping memungkinkan

analisis yang mendalam dan up-to-date terhadap dinamika ancaman disinformasi ini.

Web scraping atau juga disebut sebagai web extraction atau web harvesting (Zhao, 2017) adalah teknik untuk mendapatkan informasi dari situs web secara otomatis tanpa harus menyalinnya secara manual (A. Yani et al., 2019). Web Scraping telah diterapkan dalam berbagai bidang sebagai alat bantu untuk pengambilan data seperti analisis topik penelitian kesehatan Indonesia (Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti et al., 2017) dan juga penerapan web scraping pada marketplace Indonesia (Sembiring et al., 2020). Selain itu, penelitian mengenai penggunaan artikel berita sebagai proksi pernah dilakukan untuk menangkap volume minat di ranah publik tentang perusahaan penyedia transportasi di Amerika Serikat (Willis & Tranos, 2020). Hasil dari web scraping sendiri dapat dimanfaatkan kembali oleh sistem lain maupun dianalisis lebih lanjut (Rahmatulloh & Gunawan, 2020).

### Metode Penelitian



Gambar 1. Tahapan metode penelitian

Gambar di atas menunjukkan tahapan metode penelitian yang dilakukan. Data yang digunakan adalah fact-checking data dari situs TurnBackHoax.id, platform terpercaya yang dijalankan oleh Mafindo (Masyarakat Anti Fitnah Indonesia). Situs ini dikelola oleh organisasi Masyarakat Anti Hoax Indonesia (MAFINDO) (Tentang Kami - TurnBackHoax, 2016). Setiap berita yang terdapat pada situs ini memiliki penanda yang menerangkan apakah berita tersebut merupakan berita fakta atau hoaks, isi hoaks, fakta yang dicek oleh tim verifikator, dan kesimpulan. Data yang dikumpulkan sebanyak 1000 berita artikel yang membahas hoax yang diposting mulai Januari 2025 sampai Agustus 2025.

Penulis melakukan data scraping untuk mengumpulkan informasi komperhensif dari kumpulan artikel website tersebut. Proses web scraping dilakukan dalam code Python di platform Google Colab menggunakan library Requests, BeautifulSoup, dan Pandas. Library Requets digunakan digunakan untuk membuat permintaan HTTP ke URL tertentu dan mengembalikan responsnya. Kemudian dilakukan tahap identifikasi elemen-elemen penting seperti judul artikel, informasi hoax yang dibahas pada artikel, isi pembahasan verifikasi

hoax, halaman daftar artikel untuk mendapatkan URL semua artikel yang ingin di-scrap, dan informasi lain yang ingin diambil (Deviacita et al., 2019)

Setelah HTML mentah diambil, langkah selanjutnya adalah menguraikannya menjadi struktur yang mudah dibaca. Di sinilah BeautifulSoup berperan. library ini digunakan untuk mengekstrak, memeriksa, dan mengubah data dari pohon DOM situs web yang didasarkan pada pengembangan mesin analisis HTML dan XML (Julianno et al., n.d.). Setelah mengurai HTML menggunakan BeautifulSoup, langkah selanjutnya adalah menemukan dan mengekstrak konten tertentu dari halaman tersebut. Pada tahap ini, data masih mengandung bagian dari dokumen HTML yang tidak diperlukan, seperti tanda \, \r, dan \n. Maka dari itu, dilakukan text processing untuk mengatasi masalah tersebut menggunakan fungsi regex dari library Pandas. Fungsi ini juga digunakan untuk memisahkan teks kategori dari teks judul, sehingga kategori dapat di ekstrak ke dalam kolom yang terpisah. Selanjutnya, file hasil text processing disimpan dalam ekstensi .xlsx untuk dilakukan tahapan selanjutnya yakni filtering. Filtering dilakukan untuk memilih beberapa kolom yang diperlukan dalam analisis di tahap selanjutnya.

Analisis yang dilakukan mencakup:

1. Analisis deskriptif, dengan melakukan pendeskripsian dan penggambaran terhadap data berita hoaks yang berhasil didapat. Dalam analisis, juga diteliti mengenai kemungkinan adanya tren jumlah kejadian kriminal beriringan dengan tren jumlah kasus
2. Analisis Temporal, dengan menganalisis penyebaran hoaks dari waktu ke waktu untuk mengidentifikasi periode puncak. dan tren lainnya untuk melihat apakah ada peningkatan/penurunan signifikan di awal, tengah, atau akhir periode mulai Januari sampai Agustus 2025.
3. Analisis Kategorikal & Segmentasi pada Kategori dan Sumber Media Sosial, dengan menghitung total jumlah berita untuk setiap kategori hoaks (Salah, Belum Terbukti, Parodi, Penipuan, dan Satir) dan persentase kontribusi media sosial terhadap kategori hoaks.

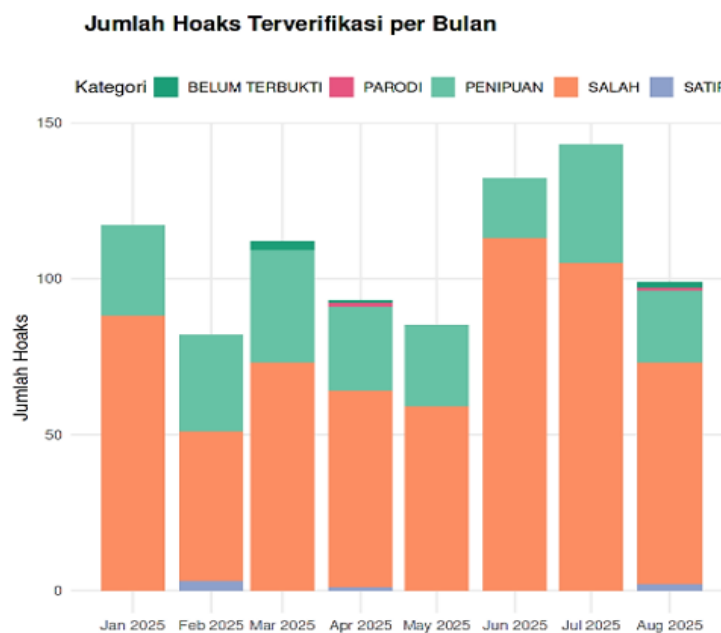
## **Hasil dan Pembahasan**

### **Hasil**

Gambar 2 menunjukkan bar chart yang memvisualisasikan total volume hoaks yang diverifikasi setiap bulan, serta komposisi kategori hoaks di dalamnya. Dari gambar di atas, dapat dilihat adanya tren fluktuatif dalam total volume hoaks yang diverifikasi dari awal tahun

hingga pertengahan tahun. Jumlah hoaks antara Januari hingga Juni 2025 cenderung stabil di kisaran 80-115 kasus per bulan, dengan sedikit penurunan di bulan Februari, April, dan Mei. Periode ini cenderung stabil. Peningkatan signifikan terjadi pada bulan Juli 2025, dimana volume hoaks tertinggi, mencapai puncaknya di sekitar 140-145 kasus. Volume hoaks menurun tajam pada Agustus 2025 (kembali ke sekitar 100 kasus), setelah puncaknya di bulan Juli.

Visualisasi gambar juga menunjukkan proporsi yang berbeda antara kategori hoaks dan hoaks yang di publikasikan di berbagai media. Kategori “Salah” adalah kontributor utama hoaks di hampir semua bulan. Proporsinya sangat besar, terutama pada paruh kedua periode. Kontributor kedua terbesar adalah “Belum Terbukti, yang berperan signifikan dalam menjaga total volume hoaks. Kategori “Parodi dan Penipuan” Secara konsisten memiliki volume yang sangat rendah, hampir tidak terlihat, dan tidak signifikan dalam total volume hoaks. Kategori “Satir” secara umum rendah, tetapi menunjukkan sedikit peningkatan di awal tahun (Januari - Maret), setelah itu menjadi minimal.



Gambar 2. Jumlah Total hoaks per Kategori Januari – Agustus 2025

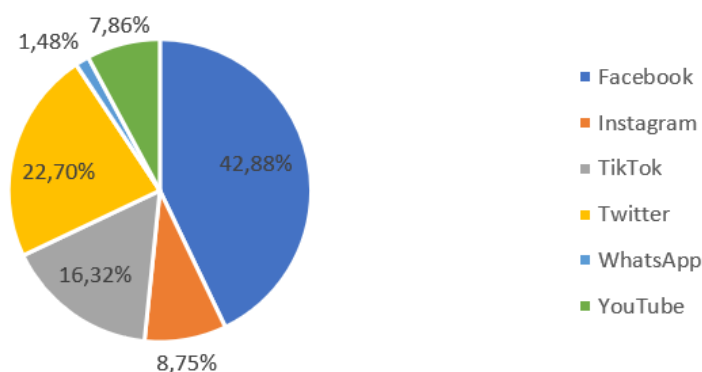
Analisis berdasarkan kategori (warna) menunjukkan dominasi yang sangat jelas dari dua kategori utama: “Salah” dan “Belum Terbukti”. Pada awal tahun, mulai dari Januari hingga Maret 2025, kategori “Salah” mendominasi volume, tetapi “Belum Terbukti” juga menyumbang porsi besar (sekitar 20-30% dari total). Pada Maret, terjadi sedikit peningkatan total volume dengan kontribusi yang seimbang dari kedua kategori besar. Volume total hoaks kedua kategori pada bulan April sampai Juni sedikit lebih rendah, dimana proporsi hoaks

“Salah” dan “Belum Terbukti” terlihat cukup berimbang.

Pada periode puncak, yang terjadi pada bulan Juli, kategori “Belum Terbukti” juga meningkat, tetapi dominasi peningkatan berada pada kategori SALAH. Kategori ini menjadi pendorong utama total volume tertinggi di bulan ini (mencapai sekitar 100-105 kasus). Meskipun volume hoaks menurun pada bulan Agustus, kategori “Salah” tetap menjadi kategori terbesar (sekitar 75-80 kasus), dan proporsi “Belum Terbukti” menurun kembali.

Untuk memberikan pemahaman lebih detail mengenai platform mana yang paling berperan dalam penyebaran hoaks berkategori “Salah” dan “Belum Terbukti”, dilakukan visualisasi lanjutan.

Kontribusi Media Sosial Kategori Hoaks "Salah"

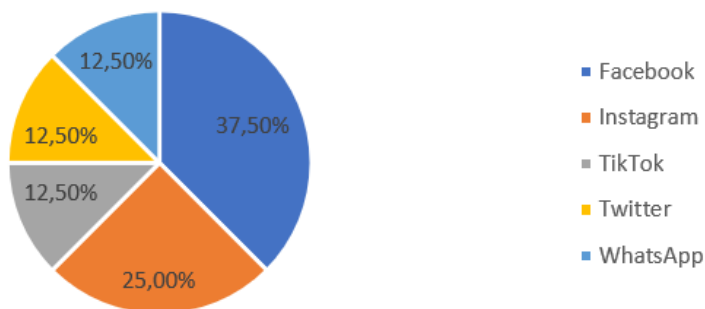


Gambar 3. Persentase Kontribusi Media Sosial terhadap Kategori Hoaks “Salah”

Gambar 3 menunjukkan sebaran platform media sosial untuk hoaks yang telah diverifikasi dan dikonfirmasi sebagai informasi yang tidak benar atau menyesatkan. Facebook adalah kontributor yang sangat dominan, menyumbang 42,88% dari hoaks "Salah". Angka ini sedikit lebih tinggi dibandingkan peran Facebook pada hoaks "Belum Terbukti" (37,50%), mengukuhkan posisinya sebagai platform utama untuk penyebaran disinformasi faktual. Twitter (sekarang sering disebut X) melonjak menjadi kontributor terbesar kedua dengan 22,70%. Peran Twitter dalam hoaks "Salah" hampir dua kali lipat dibandingkan perannya dalam hoaks "Belum Terbukti" (12,50%). Ini masuk akal karena Twitter sering menjadi platform untuk menyebarkan berita atau klaim faktual yang cepat, yang kemudian terbukti salah. TikTok menyumbang 16,32%. YouTube (7,86%) dan Instagram (8,75%) memiliki kontribusi yang relatif rendah. Secara khusus, peran Instagram turun signifikan pada hoaks "Salah" (8,75%) dibandingkan hoaks "Belum Terbukti" (25,00%). Ini mungkin menunjukkan bahwa klaim faktual yang terbukti salah lebih banyak menyebar di platform teks/berita

(Facebook, Twitter) daripada platform visual/komunikasi pribadi (Instagram, WhatsApp). WhatsApp memiliki peran paling minim, hanya 1,48%.

Kontribusi Media Sosial Kategori Hoaks "Belum Terbukti"



Gambar 4. Persentase Kontribusi Media Sosial terhadap Kategori Hoaks “Belum Terbukti”

Gambar 4 menunjukkan sebaran platform media sosial untuk hoaks yang belum dapat diverifikasi atau dibuktikan kebenarannya. Facebook adalah penyumbang terbesar dengan 37,50% dari total hoaks "Belum Terbukti". Ini menunjukkan bahwa Facebook adalah platform utama di mana informasi yang ambigu atau meragukan seringkali disebarkan atau mulai diselidiki. Instagram adalah kontributor terbesar kedua dengan 25,00%. Instagram, yang berbasis visual, mungkin menjadi saluran penting untuk menyebarkan klaim yang kurang bukti, mengandalkan gambar atau video pendek untuk dampak emosional. Tiga platform lain—TikTok, Twitter, dan WhatsApp—memiliki kontribusi yang sama persis, yaitu masing-masing 12,50%. Ini mengindikasikan bahwa ketiga platform ini memiliki peran yang setara dalam penyebaran hoaks yang masih dalam tahap keraguan. Tiga platform teratas (Facebook, Instagram, dan TikTok) menyumbang 75% dari hoaks "Belum Terbukti".

## **Pembahasan**

Dua kategori utama, yaitu "Salah" (disinformasi faktual yang terbukti keliru) dan "Belum Terbukti" (klaim yang meragukan/kurang bukti), secara konsisten menjadi kontributor utama dari seluruh hoaks terverifikasi dari Januari hingga Agustus 2025. Kategori lain (Parodi, Penipuan, Satir) memiliki volume yang sangat minim. Terdapat lonjakan volume hoaks yang signifikan pada Juli 2025, yang didorong terutama oleh peningkatan besar pada kategori "Salah".

Facebook konsisten menjadi platform utama untuk penyebaran hoaks, baik yang terbukti salah maupun yang masih diragukan. Hoaks yang bersifat meragukan/belum terbukti lebih banyak tersebar di platform visual seperti Instagram. Sebaliknya, hoaks yang bersifat faktual salah (seringkali berupa klaim berita) lebih banyak didorong oleh Twitter dan Facebook. Analisis ini sangat membantu dalam menentukan strategi mitigasi: untuk hoaks faktual, fokus perlu diarahkan pada Facebook dan Twitter; sementara untuk informasi yang meragukan, fokus juga perlu melibatkan Instagram.

## **SIMPULAN**

Berdasarkan hasil analisis, dapat disimpulkan strategi penanganan yang bersifat tersegmentasi dan bertujuan ganda, yaitu menangani volume tinggi dan menyerang titik penyebaran spesifik, yaitu Strategi Konten dan Strategi Jangka Panjang.

Strategi Konten yang dapat dilakukan meliputi pembangunan Jaringan Fact-Checker Lokal, untuk meningkatkan kecepatan dan jangkauan verifikasi hoaks, terutama untuk isu-isu lokal yang sering beredar. Data timestamp dan volume harian dapat digunakan untuk memantau kapan dan di mana hoaks mulai menyebar (Jam Sibuk Hoaks) sehingga tim mitigasi dapat merilis klarifikasi pada waktu yang paling efektif. Selain itu semua stakeholder, baik pemerintah dan masyarakat juga perlu fokus pada edukasi visual yang mengajarkan audiens dalam memverifikasi sumber sebelum disebar.

Strategi Jangka Panjang meliputi Kemitraan Platform, yang mendorong implementasi penandaan (labeling) konten hoaks yang lebih agresif dan cepat oleh media sosial, terutama di periode kritis (misalnya, saat ada isu besar) serta mendorong penurunan jangkauan (de-amplification) pada akun-akun yang terbukti secara konsisten menyebarkan hoaks.

## **DAFTAR PUSTAKA**

- Ayani, D. D., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 7(4), 257. <https://doi.org/10.26418/justin.v7i4.30930>
- Deviacita, D., A., Sasty, H., P., Muhandi, H., Profesor, J., Nawawi, D. H. H., Laut, B., Tenggara, P., Pontianak, K., & Barat, K. (2019). Implementasi Web Scraping Untuk Pengambilan Data Pada Situs Marketplace. 7(4).

- Keputusan Dirjen Penguatan Riset Dan Pengembangan Ristek Dikti, S., Sahria, Y., & Hatta Fudholi, D. (2017). Analisis Topik Penelitian Kesehatan Di Indonesia Menggunakan Metode Topic Modelling Lda (Latent Dirichlet Allocation). *I(3)*, 336–344.
- Rahmatulloh, A., & Gunawan, R. (2020). Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar. *Indonesian Journal of Information Systems*, *2(2)*, 16. <https://doi.org/10.24002/ijis.v2i2.3029>
- Sembiring, F., Yudistyal, D., Sari, D. P., Sistem Informasi, P., Pendidikan, P., Dan, S., & Informasi, T. (2020). Penerapan Teknik Scraping Python Pada Website Marketplace Indonesia. *2(1)*, 15–22.
- Setiawan, A., Santoso, L. W., & Adipranata, R. (2020). Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier. *Jurnal Infra*, *8(1)*, 146–151.
- Sonya, I. P., & Prihandoko. (2016). Analisis Web Scraping untuk Data Bencana Alam dengan Menggunakan Teknik Breadth-First Search Terhadap 3 Media Online. *Jurnal Ilmiah Informatika Komputer Universitas Gunadarma*, *21(3)*, 69–77.
- Tentang Kami - TurnBackHoax. (2016). *TurnBackHoax*. <https://turnbackhoax.id/tentang-kami/>
- Willis, George & Tranos, Emmanouil. (2021). Using ‘Big Data’ to understand the impacts of Uber on taxis in New York City. *Travel Behaviour and Society*, *22*, 94–107. 10.1016/j.tbs.2020.08.003.
- Zhao, B. (2017). Web Scraping. In *Encyclopedia Of Big Data* (pp. 1–3). Springer International Publishing. [https://doi.org/10.1007/978-3-319-32001-4\\_483-1](https://doi.org/10.1007/978-3-319-32001-4_483-1)