

Era digital menempatkan Instagram sebagai sumber data perilaku dan sentimen publik yang kaya, terutama di Indonesia dengan  $\pm 103$  juta pengguna (36,3% populasi) pada awal 2025. Ketiadaan API resmi memaksa peneliti memakai dua jalur non-formal: *reverse-engineering* API (Instagrapi) dan *browser-based scraping* (Puppeteer, Playwright). Penelitian ini (1) mengimplementasikan ketiga alat tersebut dan (2) membandingkan efisiensi CPU, RAM, *bandwidth*, serta kecepatan pengambilan data pada skenario paralel 1, 3, dan 5 *worker*.

Metode eksperimen mencatat metrik komputasi secara *real-time*. Untuk menilai signifikansi perbedaan performa, digunakan pendekatan statistik non-parametrik karena data tidak memenuhi asumsi normalitas. Uji Friedman diterapkan untuk perbandingan umum antar ketiga metode, diikuti oleh uji lanjutan Wilcoxon dengan Koreksi Bonferroni ( $p < 0,0167$ ) untuk mengidentifikasi pasangan mana yang berbeda secara spesifik. Analisis ini dilengkapi pemodelan regresi linier guna memproyeksi kebutuhan sumber daya pada volume data yang lebih besar.

Hasilnya, Instagrapi rata-rata 5–20 kali lebih hemat sumber daya dan tercepat mengekstraksi komentar, tetapi terbatas pada atribut dan cakupan. Sebaliknya, Puppeteer dan Playwright menghadirkan kumpulan data lebih lengkap dan kontekstual termasuk komentar tersembunyi, namun menuntut alokasi CPU, RAM, dan *bandwidth* jauh lebih besar. Peningkatan jumlah *worker* berskala hampir linier terhadap beban sistem.

Temuan menegaskan perlunya menyeimbangkan kelengkapan data dengan ketersediaan infrastruktur. Model regresi yang dihasilkan memudahkan perencanaan kapasitas perangkat keras, menyediakan pedoman metodologis efisien, adaptif, dan terukur bagi akademisi, industri, serta pembuat kebijakan yang ingin memanfaatkan data publik Instagram secara masif.

**Kata kunci:** *social media; web scraping; computational performance; scraping methods; data gathering.*

The digital era positions Instagram as a rich source of public behavior and sentiment data, particularly in Indonesia, which boasts approximately 103 million users (36.3% of the population) as of early 2025. The absence of an official API compels researchers to employ unofficial methods: API reverse-engineering (Instagrapi) and browser-based scraping (Puppeteer, Playwright). This study (1) implements these three tools and (2) compares their computational efficiency—in terms of CPU, RAM, bandwidth, and data retrieval speed—across parallel scenarios using 1, 3, and 5 workers.

The experimental methodology involved recording computational metrics in real-time. To assess the significance of performance differences, a non-parametric statistical approach was employed, as the data did not satisfy normality assumptions. The Friedman test was applied for an overall comparison among the three methods, followed by a post-hoc Wilcoxon test with Bonferroni correction ( $p < 0.0167$ ) to identify specific pairwise differences. This analysis was further supplemented by linear regression modeling to project resource requirements at larger data volumes.

The results show that Instagrapi is, on average, 5–20 times more resource-efficient and the fastest at extracting comments, but is limited in its data scope and attributes. Conversely, Puppeteer and Playwright deliver more comprehensive and contextually-rich datasets—including hidden comments—but demand significantly greater CPU, RAM, and bandwidth allocations. Furthermore, the increase in the number of workers was found to scale almost linearly with the system load.

These findings underscore the trade-off between data completeness and infrastructure availability. The resulting regression model facilitates hardware capacity planning, providing an efficient, adaptive, and scalable methodological guide for academia, industry, and policymakers seeking to leverage public Instagram data at scale.

**Keywords:** social media; web scraping; computational performance; scraping methods; data gathering.