

Web Scrapping & Analisis Teks: About AWS Dicoding

MINI BOOTCAMP DATA ENGINEER
– DIBIMBING

12 MEI, 2025

Tujuan Proyek

01

Mengambil konten dari halaman About AWS Dicoding

02

Melakukan web scraping yang mampu merender JavaScript

03

Membersihkan data teks hasil scraping

04

Menyajikan visualisasi dan insight dari teks yang diperoleh

AWS Back-End Academy

Program beasiswa pelatihan Back-End Developer
dilengkapi materi Gen AI dari Amazon Web Services
(AWS)

[Daftar Beasiswa](#)

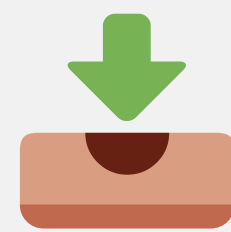


Tools & Teknologi



- Python
- requests_html – untuk scraping halaman JavaScript
- pandas – untuk pengolahan data
- wordcloud & matplotlib – untuk visualisasi





Hasil Scrapping



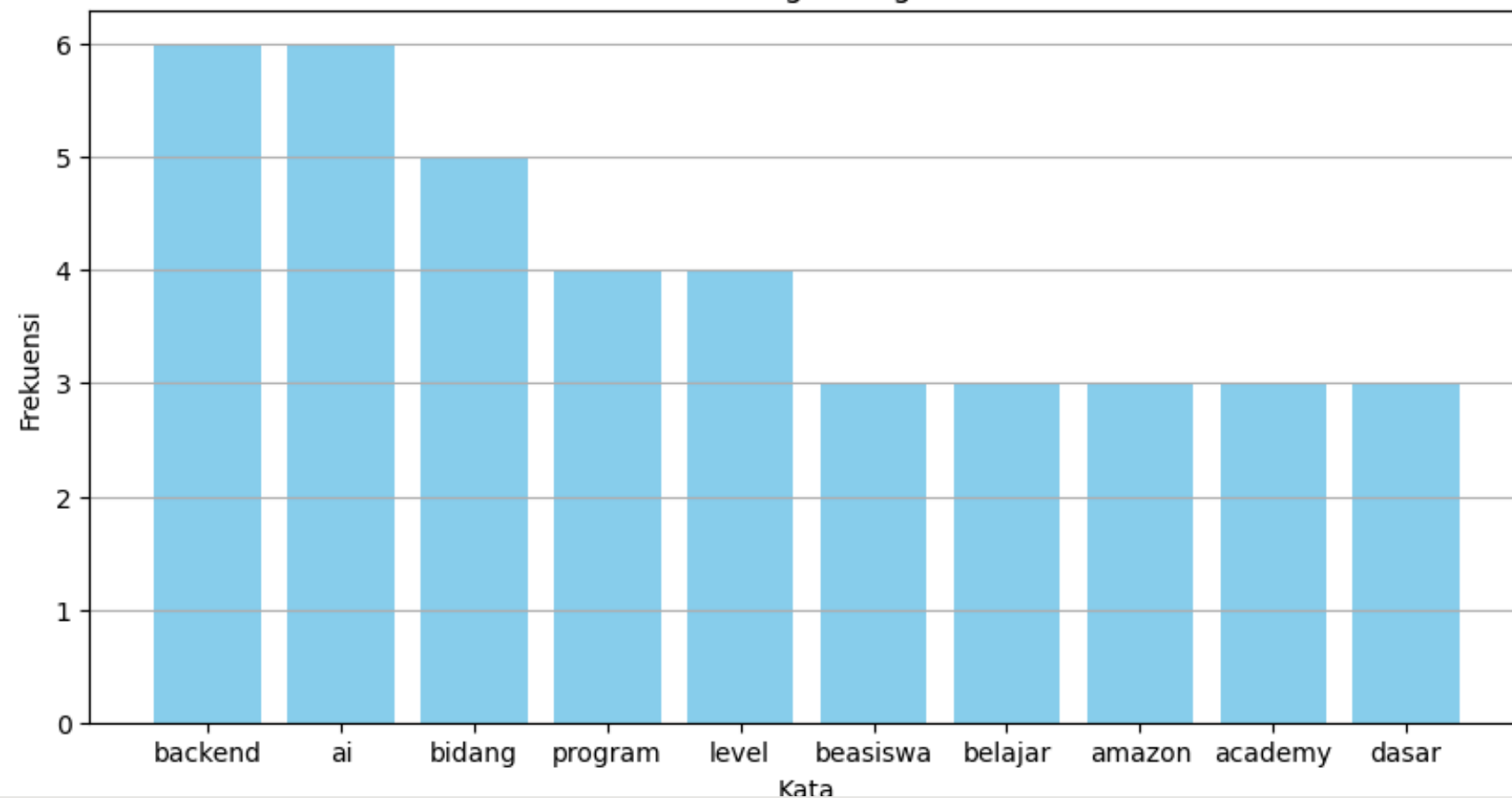
- Data berhasil diambil dari halaman <https://aws.dicoding.com/#about>
- Proses scraping dilakukan secara asynchronous menggunakan requests_html
- Data disimpan ke Excel: aws_dicoding_about_cleaned.xlsx
- Konten mencakup deskripsi umum program AWS x Dicoding

```
[INFO] Starting Chromium download.  
INFO:pyppeteer.chromium_downloader:Starting Chromium download.  
100%|██████████| 183M/183M [00:06<00:00, 27.9Mb/s]  
[INFO] Beginning extraction  
INFO:pyppeteer.chromium_downloader:Beginning extraction  
[INFO] Chromium extracted to: /root/.local/share/pyppeteer/local-chromium  
INFO:pyppeteer.chromium_downloader:Chromium extracted to: /root/.local/share/pyppeteer/local-chromium  
Scraping selesai! Disimpan sebagai aws_dicoding_about.xlsx
```



Visualisasi

10 Kata Paling Sering Muncul



- Word Cloud: menunjukkan kata-kata penting dari teks.
- Bar Chart: menunjukkan 10 kata paling sering muncul
- Kata-kata seperti: cloud, pelatihan, sertifikasi, digital, gratis, dan karier mendominasi

- Gaya bahasa pada halaman informatif dan inklusif

[illegible]

Thank you

A L I F F A U Z A N