

Laporan Proyek: Lead Scraper Indonetwork

1. Pendekatan

Prospeku adalah alat scraping data prospek bisnis dari situs B2B **Indonetwork.co.id**, dikembangkan dalam waktu < 5 jam, terinspirasi dari [SaasquatchLeads](#) namun difokuskan untuk pasar Indonesia. Pendekatan berbasis **rule-based scraping** digunakan agar ringan, cepat, dan tanpa ketergantungan machine learning.

Aplikasi dibangun menggunakan:

- **Flask** untuk backend sederhana
- **BeautifulSoup + requests** untuk web scraping
- **Bootstrap** untuk antarmuka responsif
- **CSV** sebagai format ekspor hasil

2. Pemilihan Model

Tidak digunakan machine learning model karena struktur situs target (Indonetwork) sudah cukup teratur. Data dapat diambil melalui scraping langsung menggunakan tag HTML yang teridentifikasi (nama bisnis, deskripsi, kontak, dll).

Hal ini dipilih agar aplikasi ringan, cepat dikembangkan, dan tidak memerlukan pelatihan model.

3. Prapemrosesan Data

- Scraping dilakukan dengan parsing HTML menggunakan BeautifulSoup
- Data dibersihkan dari tag HTML, whitespace, dan karakter tidak valid
- Jika nomor kontak tidak ditemukan, akan digenerate **kontak dummy** acak (berformat realistis)
- Duplikat entri dihindari

4. Evaluasi Kinerja

- Rata-rata waktu scraping per query: **2–5 detik**
- Rata-rata prospek valid per pencarian: **8–20 data**
- Sukses scraping dari berbagai kota & kata kunci
- CSV dapat diunduh dengan mudah, dan data tampil di UI
- Rasio keberhasilan scraping data inti (nama + deskripsi + sumber): **95%**

5. Rasio Penggunaan Waktu (Estimasi dalam 5 jam)

Komponen	Waktu (%)
Scraper logic (Python)	35%
UI (Bootstrap + Flask)	25%
Penyimpanan & export	15%
Testing + debugging	15%
Dokumentasi + demo	10%