# Product Requirements Document (PRD)

\*\*Product Name:\*\* Web Scraper Pro  
\*\*Version:\*\* 1.0  
\*\*Author:\*\* Dansu John  
\*\*Email:\*\* dansu.jw@gmail.com  
\*\*Date:\*\* October 2025

## 1. Overview

Web Scraper Pro is an automated web data extraction tool designed to collect, clean, and structure data from online sources. Users can define target URLs, specify data points, and export results in CSV, Excel, JSON, or database-ready formats.

## 2. Objectives

1. Automate data collection.  
2. Support dynamic sites.  
3. Provide a dashboard.  
4. Store and export data.  
5. Ensure scalability.  
6. Comply with scraping standards.

## 3. Target Users

Data Analysts, Researchers, Developers, and Businesses who need structured web data.

## 4. Core Features

- Web Scraping Engine  
- Data Extraction via XPath, CSS, Regex  
- Scheduling & Automation  
- Data Storage & Export (CSV, JSON, SQL)  
- Dashboard & UI  
- Error Handling & Compliance

## 5. Advanced Features (Future Versions)

- AI-based Auto Mapping  
- Smart Change Detection  
- Multi-Source Merging  
- Browser Simulation Mode  
- Custom Script Injection

## 6. Technical Requirements

Frontend: React + TailwindCSS  
Backend: Django or FastAPI  
Scraping Engine: Scrapy or Playwright  
Database: PostgreSQL or MongoDB  
Task Queue: Celery + Redis  
Deployment: Docker + Kubernetes

## 7. Security & Compliance

HTTPS encryption, Role-based access, Data encryption, Logging, Compliance with GDPR and robots.txt.

## 8. Metrics & KPIs

- Scrape success rate ≥ 95%  
- Job setup ≤ 3 minutes  
- Export ≤ 10s  
- API latency ≤ 500ms

## 9. User Flow Example

1. Login → 2. Add URL → 3. Select Data → 4. Export Format → 5. Schedule/Run → 6. Monitor → 7. Download

## 10. Testing & Validation

Unit, Integration, Load, and UI Testing using Playwright, Locust, and Cypress.

## 11. Timeline

Phase 1: Core Engine (3w)  
Phase 2: Dashboard & API (4w)  
Phase 3: Export (2w)  
Phase 4: Automation (3w)  
Phase 5: Testing & Deployment (2w)

## 12. Monetization

Free tier (limited jobs), Pro (unlimited + API), Enterprise (dedicated resources).

## 13. References

- Scrapy Docs  
- Playwright API  
- Celery Task Queue  
- Django REST Framework