# Data Scraping & Lead Generation SaaS – Full Documentation

## 1. Product Overview

Name (Working Title): LeadHarvest AI  
Tagline: “Fresh, Verified Leads – On Demand.”  
  
Vision:  
To become the fastest and most accurate source of verified B2B leads by combining real-time multi-source scraping, automated cleaning, geographic targeting, and email verification in one unified platform.  
  
Core Differentiator:  
Unlike database-driven competitors, LeadHarvest AI generates leads live from LinkedIn, Google Maps, Yellow Pages, and company websites — ensuring accuracy, freshness, custom location targeting, and niche segmentation.

## 2. Unique Selling Points (USP)

* Live, Real-Time Scraping → No outdated database.
* Multi-Source Data Gathering → LinkedIn + Google Maps + Yellow Pages + websites.
* Area Selection Feature → Choose exact city, state, or country to target leads.
* Integrated Email Verification → Bounce-free outreach.
* Industry-Specific Lead Packages → Highly targeted lists.
* Flexible Pricing → Pay-per-lead or subscription.
* API & CRM Integration → Direct pipeline to client workflows.
* Compliance-First → GDPR & CAN-SPAM adherence.

## 3. Competitor Analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Feature | LeadHarvest AI | Apollo.io | ZoomInfo | Hunter.io | Lusha | Seamless.AI |
| Real-time scraping | ✅ | ❌ | ❌ | ❌ | ❌ | ❌ |
| Multi-source data | ✅ | ❌ | ❌ | ❌ | ❌ | ❌ |
| Email verification | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ |
| Area selection | ✅ | ⚠ Partial | ⚠ Partial | ❌ | ❌ | ❌ |
| Industry targeting | ✅ | ⚠ Partial | ⚠ Partial | ❌ | ❌ | ❌ |
| Pay-per-lead option | ✅ | ❌ | ❌ | ❌ | ❌ | ❌ |
| API access | ✅ | ✅ | ✅ | ❌ | ❌ | ✅ |
| Compliance-first | ✅ | ⚠ Partial | ⚠ Partial | ✅ | ❌ | ❌ |

## 4. Tech Stack

Backend: Python (FastAPI for APIs)  
Scraping: Scrapy, BeautifulSoup, Playwright/Selenium  
Email Verification: SMTP checks, DNS lookups, Hunter/NeverBounce API  
Database: PostgreSQL (lead storage), Redis (task queue)  
Frontend: React.js (user dashboard)  
Hosting: AWS EC2 + RDS / DigitalOcean  
Proxy Rotation: Bright Data / ScraperAPI  
Security: CAPTCHA bypass, IP rotation, request throttling

## 5. Key Features

* Lead Finder – Search by industry, location, role, and size.
* Area Selection – Choose exact city, state, or country for targeted results.
* Live Data Scraping – Fetch latest leads from multiple sources instantly.
* Data Cleaning & Deduplication – Normalize and standardize.
* Email Verification – Multi-step checks to ensure delivery.
* CRM Integration – One-click export to HubSpot, Zoho, Salesforce.
* Scheduled Updates – Auto-refresh lead lists weekly/monthly.
* Download Options – CSV, Excel, JSON, or API.

## 6. Monetization Model

Pay-Per-Lead: $0.20–$0.50 per verified lead.  
Higher price for hard-to-find niches.

### Subscription Plans

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Plan Name | Monthly Price | Leads Included | Extra Leads Rate | Area Selection | Email Verification |
| Starter | $29 | 300 | $0.20 per lead | ✅ | ✅ |
| Pro | $59 | 1,200 | $0.18 per lead | ✅ | ✅ |
| Agency | $149 | 5,000 | $0.15 per lead | ✅ | ✅ |
| Enterprise | Custom | Unlimited | Negotiable | ✅ | ✅ |

## 7. Development Roadmap

* Phase 1 – Core Engine (Weeks 1–2): Build scraping modules, implement proxy rotation & CAPTCHA handling.
* Phase 2 – Verification & Cleaning (Weeks 3–4): Integrate email verification API, add deduplication.
* Phase 3 – Dashboard & Area Selection (Weeks 5–6): Create dashboard with filters & map-based picker.
* Phase 4 – Payment & Plans (Weeks 7–8): Stripe integration, lead credit system.
* Phase 5 – Marketing Launch (Week 9 onwards): SEO, LinkedIn ads, free trial launch.

## 8. Marketing Strategy

* SEO: Target 'buy verified leads for [industry] in [city/state]' keywords.
* LinkedIn Ads: Target B2B marketers, recruiters, and agencies.
* Affiliate Program: Let freelancers resell leads.
* Content Marketing: YouTube + LinkedIn posts on lead generation.
* CRM Partnerships: Cross-promotion with CRM platforms.

## 9. Compliance & Risk Mitigation

• Only scrape public business data.  
• GDPR: Provide opt-out and clear privacy policy.  
• CAN-SPAM: Only verified business emails.  
• IP rotation to avoid blocking.

## 10. Example Workflow

1. User selects:  
 - Industry: “Dentists”  
 - Location: “New York, USA” (via area selection map)  
2. System scrapes Google Maps + LinkedIn company pages + Yellow Pages.  
3. Data cleaned & deduplicated.  
4. Emails verified.  
5. Leads ready for download or CRM export.