Lead Efficiency Checking & Scoring System – Project Report

- Mohit Laddha

Repo: [Github]

Objective : To build a smart lead scoring pipeline that ranks B2B leads based on key quality indicator : title match, email validity, and website credibility , so that teams can prioritize outreach more effectively ,users see high-quality leads at the top and we can prioritise better lead generation tools.

Approach & Workflow : Our system processes a CSV of raw leads and enriches it with intelligent signals:

1. Title Relevance (10%)
   - Uses DeepSeek-V3 via Together AI for natural language scoring.
   - Prompted to rate how close a given job title is to the target role.
   - Output: Integer score (0–100).
2. Email Verification (50%)
   - Uses NeverBounce API to classify emails as valid, disposable, catchall, or invalid.
   - Output: Email status + suggested correction if available.
3. Website Authenticity (40%)
   - Scrapes domain content with requests + BeautifulSoup.
   - Classifies using DeepSeek-V3 into: real_business, placeholder, junk_or_unclear, or unreachable.
4. Scoring Logic
   - Final lead_score = weighted sum of the above.
   - Output CSV contains: email_status, website_status, title_score, lead_score.
   - Sorted output available for decision-making.

Model & Tools Used :

- LLM Model: deepseek-ai/DeepSeek-V3 from Together.ai
- APIs: NeverBounce (Email validation)
- Python Libraries: pandas, requests, beautifulsoup4, dotenv, together

Outputs :

- leads_output.csv: Full enriched data
- leads_clean_sorted.csv: Cleaned + sorted view for top leads

Summary : This project enables businesses to rank and filter B2B leads with precision using intelligent signals extracted from title relevance, email health, and website credibility. It helps sales and marketing teams focus on high-value leads, improving outreach efficiency. The solution is modular, extendable, and aligned with real-world SaaS tools.