Project Report: AI-Powered Lead Intelligence Suite

Candidate: **Kartik Dinesh Sangle**

RE: **Caprae Capital Intern Interview Pre-Work**

1. **Problem Analysis & Opportunity**

The provided reference application, saasquatchleads.com, successfully identifies potential leads. However, its primary limitation is its output: a high-volume, unsorted list that creates an operational bottleneck. A founder must manually sift through hundreds of entries to find a handful of high-potential targets.

The opportunity was to build a system that transforms this raw data into **decisive, actionable intelligence**. My goal was to create a suite of tools that not only identifies *who* to contact but, more importantly, *when* to contact them, providing true "horsepower" to a founder-led sales team.

To achieve this, I developed a two-stage solution:

- **Stage 1: Lead Scoring Engine:** To solve the initial problem of prioritization.
- **Stage 2: Growth Signal Monitoring:** To add a layer of proactive, time-sensitive intelligence.

1. Architectural Approach: A Two-Stage Intelligence System

Stage 1: Learning-to-Rank (LTR) for Prioritization

To address the core challenge of "who to contact," I built a full-stack application centered around a Learning-to-Rank (LTR) machine learning model. I deliberately avoided a simple, rule-based system in favor of this more sophisticated approach for three key reasons:

1. **Scalability:** A rule-based system is brittle. An LTR model, powered by Gradient Boosting (XGBoost), learns complex, non-linear relationships from data, making it more robust.
2. **Holistic Analysis:** The model evaluates leads based on a multi-dimensional "fit." It can learn, for instance, that high revenue is most valuable when paired with a mid-range employee count, a nuance simple rules miss.

3. **Future-Proofing:** The architecture is designed to be retrained on a portfolio company's actual sales data (won vs. lost deals). This creates a self-improving asset, directly embodying the Caprae mission of using technology to unlock long-term value.

Stage 2: Proactive Signal Monitoring for Timing

After solving the "who," I focused on the "when." I engineered a second module to detect **buying signals**—specifically, active job postings for key roles. A company hiring a "VP of Sales" is not a cold lead; they are a warm prospect with a validated need and budget.

This feature transforms the tool from a reactive database to a proactive intelligence engine, providing a critical competitive edge.

1. Technical Workflow & Justification

- **Data Acquisition & LTR:** A user's query (industry/location) hits the Google Custom Search API. The resulting websites are scraped concurrently with BeautifulSoup. This data is then fed into a pre-trained xgboost.XGBRanker model to generate a prioritized list of leads. I chose XGBoost for its industry-proven performance and native support for rank:pairwise objectives.
- **Signal Detection:** For the second module, the system takes a list of high-priority company URLs. It uses **Selenium** to automate a real browser, performing targeted searches on platforms like LinkedIn Jobs. This robust approach avoids simple IP blocks and accurately determines if a company is actively hiring for a "trigger" role, providing a clear "Yes/No" signal.

1. Business Impact & Alignment with Caprae's Vision

This project is a direct reflection of a founder-operator mindset: it's a suite of tools built not for the sake of technology, but to solve core business problems and create a competitive advantage.

- **For a Founder:** It saves dozens of hours of manual labor. Stage 1 tells them where to focus their energy. Stage 2 tells them the perfect moment to strike, maximizing the impact of their outreach.
- **For Caprae:** It serves as a blueprint for a scalable, data-driven sales process for portfolio companies. It is a tangible example of how practical AI can be leveraged post-acquisition to immediately enhance operational efficiency and drive top-line revenue growth.

By integrating intelligent ranking with proactive signal monitoring, this suite demonstrates a commitment to creating systems that deliver not just data, but decisive, actionable insights.