AI Handbook - Technical Report

Submitted by - R Sai Shivani

Approach Overview

I implemented a hybrid rule-based + AI-powered system to enrich and score B2B leads. Our architecture supports:

- Rule-based scoring logic (0–100 scale)
- Azure OpenAI GPT-based insight generation
- Scalable chunked CSV processing
- REST API backend (FastAPI) with a React frontend

Design principles include interpretability, speed, and actionable insights.

Data Preprocessing

Transformations Applied:

- company_age = 2025 - founded
- size_bucket mapped to 1–8 scale from size_range
- Normalizations -
    - norm_size = (size_bucket - min) / (max - min + ε)
    - norm_age = (max - age) / (max - min + ε)
- Boost calculations -
    - size_boost = norm_size^1.5
    - age_boost = norm_age^1.8
- Flags -
    - maturity_bonus: age > 20 and size ≥ 7
    - growth_bonus: current_emp < 70% of total_emp

Model Selection

1. **Lead Scoring Engine -** Rule-based formula
   Scoring logic: lead_score = 0.5 * size_boost + 0.4 * age_boost + 0.05 * maturity_bonus + 0.05 * growth_bonus
2. **Insight Generator -** Azure OpenAI GPT-4
   Prompt designed to explain lead score using only provided data and provide a structured output.

Performance & Benefits Summary

1. **High Throughput & Low Latency -**
    - Processes ~10,000 rows per minute on a 4 vCPU setup
    - Handles input CSVs up to 100MB
2. **Cost-Effective & Scalable -**
    - <3% JSON error rate in AI outputs
    - No training data required
    - Supports parallel processing with low memory usage
3. **Transparent & Business-Aligned**
    - Easy tuning through rule update
    - Direct mapping to business priorities