Inductiv AI Engineer Take-Home Assignment

Loan Application Intelligence System

Objective

Build an intelligent system that demonstrates your expertise in both **classical machine learning** and **modern AI/LLM capabilities** using loan application data.

Overview

Create a system that:

1. Analyzes loan application data using ML-based classification and prediction
2. Builds an intelligent chatbot loan assistant that answers questions about loan data
3. Evaluates the quality and accuracy of AI-generated responses

Dataset

You can use any freely available loan application data.
For example,
https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv

Requirements

Part 1: Data Processing & ML Classification

- Load and analyze the provided loan application CSV dataset
- Build a classifier to predict loan approval status
- Achieve **F1-score ≥ 0.75** on test set
- Use traditional ML algorithms (your choice)
- Implement proper train/test split and validation
- Document feature engineering choices and model performance

Part 2: Intelligent Loan Assistant Chatbot

- Make loan application data queryable by the chatbot
- Build a REST API that allows users to ask questions about loans and applicants
- Use the trained ML model to provide loan approval predictions
- **Implement your own retrieval logic** (NO LangChain/LlamaIndex for retrieval implementation)
- Use free/open-source LLMs (Ollama or Hugging Face)
- Return responses with source information and confidence metrics

Part 3: AI Evaluation & Metrics

- Create an evaluation dataset with 10-15 test questions covering diverse query types
- Include factual, aggregation, comparative, prediction, and unanswerable questions
- Design and implement appropriate evaluation metrics for:
    - Retrieval quality
    - Answer generation quality
    - End-to-end system performance
- Provide analysis (0.5-1 page) of system performance, failure modes, and improvements

Bonus Points 🌟

Graph-Based Knowledge Representation

- Build a knowledge graph from loan application data
- Implement graph-based retrieval for the chatbot
- Compare performance against your base implementation

Domain Ontology Development

- Develop a formal ontology for the loan application domain
- Document your ontology design and rationale
- Demonstrate how it improves system reasoning

Deliverables

1. Code

Choose **one** format:

- **Option A**: Jupyter Notebook with clear sections and explanations
- **Option B**: Structured Python project

2. Documentation

- **README.md**: Setup, usage, architecture, and design choices
- **EVALUATION_REPORT.md**: ML performance, evaluation metrics, analysis, and methodology

3. Demo Materials

- Example queries and responses
- Visualizations (optional but encouraged)

4. Artifacts

- Trained ML model
- Evaluation results
- Test questions with ground truth
- requirements.txt

Constraints

- **Language**: Python preferred (notify us if using another language)
- **Free tools only**: Use open-source models and libraries (no paid APIs like GPT-4, Claude API)
- **No frameworks for retrieval**: LangChain/LlamaIndex can be used for LLM prompting only, NOT for retrieval implementation

Timeframe

- **Deadline**: Complete within **7 days** of receiving this assignment
- **Expected Effort**:
    - Core requirements: 3-4 hours
- Notify us if you cannot complete within the timeframe

Important Notes

⚠️ **Critical Requirements:**

1. Implement retrieval logic yourself

2. Design comprehensive evaluation metrics
3. Provide measurable results with concrete numbers
4. Production-quality code with documentation

💡 **What We're Evaluating:**

- ML evaluation methodology and rigor
- Understanding of intelligent chatbot system design
- Ability to measure and analyze AI system performance
- Code quality and documentation
- Critical analysis of limitations
- (Bonus) Advanced AI/ML concepts

Questions?

Reach out if anything needs clarification at saurav@inductiv.co. Clear communication is valued!

Good luck! 🚀