Project Scope and Deliverables

Background and Motivation

The global rice market represents a critical component of food security and economic stability across Asia, yet Small and Medium Enterprises (SMEs) in this sector face significant operational challenges. Traditional ERP systems remain prohibitively expensive and complex for these small businesses. At the same time, the lack of intelligent data processing capabilities prevents them from making informed decisions about inventory, pricing, and market positioning.

Our team brings together complementary expertise in machine learning operations, distributed systems architecture, and domain knowledge of agricultural supply chains. Through previous coursework in distributed systems and natural language processing, we've identified the potential for AI-powered microservices to democratize enterprise capabilities for underserved markets. The rice market presents an ideal test case, combining structured transactional data with unstructured documents and time-series pricing patterns that showcase the full spectrum of MLOps challenges.

This project addresses three critical pain points we've observed through our research: First, business users struggle with SQL syntax when querying inventory and sales data. Second, valuable insights are often locked in unstructured documents, such as PDFs and Excel reports. Third, price volatility makes inventory planning extremely difficult without sophisticated forecasting capabilities.

Scope and Objectives

Our project will develop a production-ready microservices architecture that transforms how SMEs interact with their business data through three core AI capabilities:

Primary Objective: Build a scalable, secure AI ML system (for the scope of this project, we build and deploy AI/ML capabilities with a frontend UI for interactions with these 3 AI/ML backends) that enables natural language interaction with business data, intelligent document processing, and predictive analytics for rice market operations.

Specific Goals:

1. Implement a Natural Language to SQL service that translates conversational queries into secure database operations with sub-second response times
2. Deploy a Retrieval-Augmented Generation system for summarizing and extracting insights from business documents

3. Create a time-series forecasting engine providing 6-month price predictions with explainable confidence intervals

The scope explicitly includes containerized microservices, ML model training and deployment pipelines, comprehensive security layers, and production-grade monitoring. We will demonstrate these capabilities through a functional web application that rice traders and warehouse managers can use for daily operations.

Data Sources and Strategy

Dataset: Rice Market Pricing and Inventory (not yet, need to scout from public data sources)

- **Source:** public data source on the web for SMEs rice markets
- **Description:** Time-series pricing data spanning 5 years across multiple rice varieties
- **Key Attributes:** e.g., date, variety, grade, origin location, price per metric ton, volume traded, warehouse location, moisture content, currency exchange rates
- **Relevance:** This dataset is used in price forecasting in the Time-series Forecasts function

Excel database: daily Excel operations files (already collected from a rice mill processing company, but need to pre-process to eliminate personal identification information)

- **Source:** production, warehouse, contracts, documents from a rice mill processing factory
- **Description:** 11,000+ rows of data in Excel format representing typical rice mill daily operations management
- **Relevance:** used for NL+SQL Service and RAG to extract and summarize business-critical information