# Advanced Topics in Machine Learning – Initial Project Proposal

**Team Members:** Kainat Arshad (253669), Kanika Vashisht (254074), Muhammad Abdullah (250261), Sufia Fathima (251151), Sonika Sree Yathirajam (244443)

**Selected Use-Case:** Use Case 1 – Few-shot and Zero-shot Learning paradigms using LLMs

**Overview:** The project aim is explore how different document vectorization techniques (TF-IDF, Doc2Vec, and LLM-generated summaries) impact classification performance, particularly in low-resource learning settings. We will also evaluate and compare explanation methods from both classical XAI tools and local LLM outputs.

## High-Level ML Pipeline

**0. LLM Setup**

- Use Ollama to run Gemini 3B (1B or 4B) on local machines (CPU-only).
- Interact via API calls (e.g., Python `requests` or `ollama-python` SDK).

**1. Data Preprocessing**

- Clean and tokenize documents.
- Generate 3 representations per document:(TF-IDF vectors (via `scikit-learn`), Doc2Vec embeddings (via `gensim`), LLM summaries (Gemini 3B via Ollama →embedded using `Sentence-BERT`)

**2. Model Training and Classification**

- Train and test three classifiers:(SVM (Scikit-learn SVC), Neural Networks (MLPClassifier), Decision Trees (Random Forest / XGBoost))

**3. Evaluation**

- Use metrics: Accuracy, Precision, Recall, F1-score.
- Compare performance across document representations and datasets.

**4. Explainability (XAI)**

- Apply LIME and SHAP for classifier interpretation.
- Query Gemini 3B (locally) for natural language reasoning.
- Compare and contrast explanations from XAI vs. LLM.

**5. Comparative Analysis**

- Evaluate the effectiveness of few-shot and zero-shot learning for classification in limited data scenarios
- Assess agreement/disagreement between algorithmic XAI and LLM reasoning.
- Document insights and challenges with specific examples from the AgNews dataset.