Proposal

Title: AI-Powered Content Analysis and Recommendation

Team: Code Crafters

Hackathon: CODRELATE-2025

Track: AI-Powered Content Analysis and Recommendation

Problem Statement:

o build an efficient predictive model using Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), or Large Language Models (LLM). The goal is to develop a well-optimized solution that is accurate, computationally efficient, interpretable, and user-friendly with an innovative frontend for visualization.

Motivation:

Manual tagging is inefficient and inconsistent. An AI-powered pipeline enhances user experience, search, and content recommendations.

Dataset:

- 192k articles from news platforms
- Cleaned and processed using NLP techniques
- Multi-label nature of tags

Methodology:

- Preprocessing using NLTK and custom cleaning
- TF-IDF vectorization
- Logistic Regression with OneVsRestClassifier for multi-label classification
- Frontend with Gradio for interactivity

Tech Stack:

- Python
- Scikit-learn, Pandas, NLTK
- Gradio

Innovation:

- Lightweight, scalable tag prediction engine
- Gradio UI for demo and practical use
- Explainability and confidence scores for predictions

Evaluation:

- Metrics: Hamming loss (~0.08), Precision, Recall, F1-score (~0.79 avg)

GitHub Repo:

- https://github.com/ShaikJasmin11/CodeCrafters_Round2.git