Group Members:

Muhammad Ashir 2023385

Muqeem Ahmed 203561

Project Proposal: Advanced Fake News Detection and Analysis Using ML, DL, and NLP

1. Project Title

Advanced Fake News Detection and Analysis Using Machine Learning, Deep Learning, and NLP

---

2. Problem Statement

Fake news spreads rapidly on social media, affecting public opinion and decision-making. Manual verification is slow and unreliable. An intelligent system is needed to automatically detect fake news, analyze sentiment, and provide contextual explanations.

---

3. Motivation

- Social media platforms are flooded with misinformation.
- Early detection helps prevent misinformation from spreading.
- Combining ML, DL, and NLP ensures more accurate detection and better user understanding.
- An interactive dashboard makes the system practical and user-friendly.

---

4. Dataset Source

- **Primary Dataset:** Fake and Real News Dataset
  - Contains 20,000+ news articles labeled as real or fake.
  - Fields: `title`, `text`, `label`.
- **Optional:**
  - LIAR Dataset for fine-grained labeling.
  - Sentiment or news category datasets for additional features.

---

5. Proposed Methodology

1. **Data Preprocessing:**
   - Clean text, remove stopwords, tokenize, and convert to embeddings.

2. **ML Model:**
    - TF-IDF + Random Forest / Logistic Regression / XGBoost for baseline classification.
3. **DL Model:**
    - Fine-tuned BERT or RoBERTa for advanced text classification.
4. **NLP / LLM Module:**
    - Generate summaries or explanations using GPT or T5 based on model predictions.
5. **Dashboard / UI:**
    - Streamlit or Gradio to allow users to input news articles and see:
        - Fake/real classification
        - Sentiment analysis
        - Summaries / explanations
    - Optional: visualizations like word clouds or bar charts.

---

6. Expected Outcomes

- Accurate detection of fake news with probability/confidence scores.
- Sentiment analysis for better understanding of news tone.
- Summaries or explanations highlighting misleading phrases.
- Interactive, modular dashboard for user engagement.

---

7. Tools & Libraries

- **Programming Language:** Python
- **ML/DL:** Scikit-learn, TensorFlow, PyTorch
- **NLP / LLM:** HuggingFace Transformers (BERT, GPT, T5)
- **Dashboard:** Streamlit or Gradio
- **Data Processing:** Pandas, NumPy, NLTK, SpaCy