**Slide 1: Title Slide**

* **Title**: Adverse News Classification: Approaches, Pros, Cons, and Recommendation
* **Subtitle**: Leveraging Machine Learning and LLMs for Financial Crime Detection
* **Details**: Presented by [Your Name], Date: April 06, 2025

**Slide 2: Introduction**

* **Content**:
  + Objective: Classify news articles (title, body, date) as adverse news (e.g., Money Laundering, Fraud, Sanctions Violations).
  + Challenge: No labeled data available.
  + Goal: Explore viable approaches and recommend one solution.
* **Visual**: Icon of a newspaper or magnifying glass.

**Slide 3: Approach 1 - Topic Modeling (Unsupervised Learning)**

* **Description**: Use LDA or NMF to discover latent topics in news articles.
* **How It Works**: Preprocess text → Extract features (TF-IDF/embeddings) → Cluster topics → Map to adverse categories.
* **Pros**:
  + No labeled data needed.
  + Good for exploratory analysis.
* **Cons**:
  + Topics may not align perfectly with categories.
  + Requires manual interpretation.
* **Visual**: Diagram of text → topics → categories.

**Slide 4: Approach 2 - Zero-Shot Classification with LLMs**

* **Description**: Use pre-trained LLMs (e.g., BART, T5) to classify articles via prompts.
* **How It Works**: Prompt LLM with article text and categories → Parse output.
* **Pros**:
  + No training data required.
  + Quick to implement.
  + Flexible category definitions.
* **Cons**:
  + Less precise without fine-tuning.
  + API costs at scale.
* **Visual**: Example prompt and output (e.g., “Sanctions Violations”).

**Slide 5: Approach 3 - Clustering with Embeddings**

* **Description**: Cluster articles using embeddings (e.g., BERT) and algorithms like K-Means.
* **How It Works**: Generate embeddings → Cluster articles → Analyze clusters for adverse news.
* **Pros**:
  + Captures semantic similarity.
  + No labeled data needed.
* **Cons**:
  + Requires manual cluster labeling.
  + Sensitive to clustering parameters.
* **Visual**: Scatter plot of clustered articles.

**Slide 6: Approach 4 - Hybrid (Unsupervised + Manual Refinement)**

* **Description**: Start with unsupervised grouping, label a subset, then train a classifier.
* **How It Works**: Cluster/topic model → Manually label sample → Train supervised model.
* **Pros**:
  + Balances scalability and accuracy.
  + Evolves into a supervised system.
* **Cons**:
  + Initial manual effort required.
  + More complex to implement.
* **Visual**: Flowchart (Unsupervised → Manual → Supervised).

**Slide 7: Approach 5 - Keyword-Based Heuristics + LLM Validation**

* **Description**: Filter articles with keywords, refine with LLM classification.
* **How It Works**: Define keywords (e.g., “bribe”) → Filter articles → LLM confirms category.
* **Pros**:
  + Simple and fast to start.
  + Leverages domain knowledge.
* **Cons**:
  + Misses nuanced cases.
  + Keyword list maintenance needed.
* **Visual**: Table of keywords and categories.

**Slide 8: Comparison Table**

* **Content**:

| **Approach** | **No Labeled Data** | **Speed** | **Accuracy** | **Complexity** |
| --- | --- | --- | --- | --- |
| Topic Modeling | Yes | Medium | Medium | Medium |
| Zero-Shot LLM | Yes | Fast | Medium-High | Low |
| Clustering | Yes | Medium | Medium | Medium |
| Hybrid | Partial | Slow | High | High |
| Keyword + LLM | Yes | Fast | Low-Medium | Low |

* **Visual**: Table format.

**Slide 9: Recommended Solution - Zero-Shot Classification with LLMs**

* **Why Recommended**:
  + **No labeled data required**: Ideal for your current dataset.
  + **Fast implementation**: Get results quickly with pre-trained models.
  + **Flexible and scalable**: Adjust categories via prompts, process large datasets.
  + **Iterative improvement**: Validate results and transition to supervised if labels are added later.
* **Implementation**:
  + Use Hugging Face’s zero-shot-classification pipeline (e.g., BART).
  + Example: “Classify: Money Laundering, Fraud, None” → Apply to title + body.
* **Next Steps**: Spot-check results, refine prompts, consider fine-tuning if labels become available.
* **Visual**: Code snippet or workflow (Text → LLM → Category).

**Slide 10: Conclusion**

* **Content**:
  + Multiple viable approaches exist for adverse news classification without labels.
  + Zero-Shot LLM offers the best balance of speed, ease, and effectiveness for your case.
  + Future potential: Build labeled data over time for higher accuracy.
* **Call to Action**: Start with Zero-Shot, evaluate, and adapt based on results.
* **Visual**: Checkmark or roadmap icon.