Detecting Fake News from Headlines

A Comparative Analysis of Machine Learning and Transformer Models

Presenter: Vinicius Costa & Anirudh Unni

Date: September 4, 2025

The Problem & Our Goal

- The Problem: Misinformation is a significant challenge. Can we automatically identify fake news articles based only on their headlines?
- Our Goal: To build and evaluate different models to find the most accurate and reliable method for this task.
- Input: A news headline (e.g., "President Announces New Policy").
- Output: A classification of Real (0) or Fake (1).

Our Approach

- We tested two main strategies:
- Traditional Machine Learning: Trained classic models from scratch on our dataset.
- → Relies on statistical features of the text.

- Pre-trained Transformer (BERT): Used a massive, pre-trained language model for detecting real & fake news from articles and adapted it.
- → Leverages deep understanding of English language.

Part 1 - Traditional Models

Methods

- Feature Extraction: Converted headlines into numerical data using TF-IDF
- Models Trained:
  - Logistic Regression
  - Naive Bayes
  - Random Forest

- Best Performer: The Ensemble Model
  - Combines predictions with majority vote

Results of Traditional Models

- The Ensemble model was the clear winner:
  - Ensemble Model: 94.0%
  - Logistic Regression: 93.0%
  - Random Forest: 92.0%
  - Naive Bayes: 92.0%

- Conclusion: Combining models using Ensemble provides the most robust performance

Part 2 - Using a Pre-trained Transformer

- We tested a pre-trained model (Fake-News-Bert-Detect) without modifications → 'Zero-Shot' evaluation

- Question: Can a general-purpose language model outperform our specialized Ensemble model?

A Surprising Result

- Our Ensemble Model: 94.0% Accuracy
- Pre-trained BERT (Zero-Shot): 65.3% Accuracy

- Why? The pre-trained model's knowledge is broad but not specialized
- It needs adaptation to our specific headline data

Part 2 - Using a Pre-trained Transformer with fine-tuning

Key Findings

- **The True Baseline was 0.65:** The pre-trained model's starting accuracy on our validation set was 0.65
- **No Improvement After Fine-Tuning:**
- Used Auto-tokenizer to tokenize training data for pre-trained model
- After fine-tuning for three epochs, the model's accuracy on the same validation set remained at 0.65

Interpretation

Fine-tuning process was simply ineffective with the current settings. This lack of improvement suggests:

- **Insufficient Training:** More epochs may be needed for the model to learn from the data.
- **Sub-optimal Hyperparameters:** The learning rate and other training arguments may need to be adjusted.
- **Model-Data Mismatch:** The model, possibly trained on full articles, may struggle to adapt to the specific style of headlines without more extensive training.

Final Results & Conclusion

- Ensemble Model: 94.0%
- Pre-trained BERT (Zero-Shot): 65.3%
- Fine-Tuned BERT: 65.3% Accuracy

Key Takeaways

- Custom Ensemble models are strong baselines.
- Selecting the right pre-trained model and fine-tuning them for the specific task is very important.

THANK YOU