Detecting Political Deception Using Deep Learning Models

Overview:

This project aimed to automatically detect deceptive political statements using natural language processing (NLP) and deep learning techniques. The research was motivated by the growing need to identify misinformation in political discourse, especially with the rise of AI-generated content and social media influence.

Dataset:

We used a binary-labeled version of the **PolitiFact** dataset, which contains over 14,000 political claims. Labels like "Half True", "Pants on Fire", etc., were simplified into binary classes: *True* vs. *False*.

Models Used:

Four models were implemented and compared:

- **BiLSTM**
- **BERT (bert-base-uncased)**
- **DistilBERT (distilbert-base-uncased)**
- **RoBERTa (roberta-base)**

Each model was fine-tuned with specific configurations, trained using cross-entropy loss, and evaluated on accuracy, precision, recall, F1-score, and ROC-AUC.

Key Results:

Model	F1-Score	Accuracy	ROC-AUC
BiLSTM	0.51	63.8%	0.665
BERT	0.59	67.3%	0.7127
DistilBERT	0.29	47.1%	0.6493
RoBERTa	**0.69**	**69.1%**	**0.758**

RoBERTa showed superior performance due to better pretraining techniques and dynamic masking. BiLSTM was surprisingly competitive, indicating that simpler models can still be effective when tuned properly.

Conclusion:

Transformer-based models, especially RoBERTa, are powerful tools for detecting deception in political texts. However, traditional models like BiLSTM still offer value, particularly in low-resource environments. This research lays the groundwork for more scalable and accurate misinformation detection systems.