INFOTRUST

~Deep Learning & NLP for News Credibility Analysis

A Multi-Dataset Framework for Misinformation & Trustworthiness Evaluation

Team Members:

Lingala Anwesh [B200462]
Santhapuri Nishanth [B200628]
Pittala Mahesh [B200240]

Guideness By:

Assistant Proffesor
Mrs. Govaradhani Kummari

The Challenge of Digital Trust

The proliferation of digital **platforms has unfortunately led** to spread of **fake news and misinformation. This erodes public trust, influences critical decisions, and can have far-reaching societal consequences**. Our project addresses this urgent need for a reliable mechanism to assess news credibility.

Our proposed solution is a **multi-dataset, Deep Learning, and Natural Language Processing (NLP) framework** designed to analyze and evaluate news credibility. This comprehensive approach provides a robust defense against the tide of misinformation.

The applications of this framework are vast, **benefiting journalism, social media platforms, and policy-making bodies** by providing tools for rapid and **accurate content evaluation.**

Leveraging Diverse Datasets

To build a robust and generalizable model, we consolidate information from multiple, distinct datasets, creating a comprehensive foundation for credibility analysis.

FakeNewsNet

Comprises **news articles** from GossipCop and PolitiFact, including news titles, full articles, and associated tweet interactions. It offers a rich source for analyzing real-world news propagation.

LIAR Dataset

Features short political statements with fine-grained credibility labels (e.g., true, false, pants-on-fire), primarily focused on political discourse.

Kaggle Fake News Dataset

Contains a large collection of articles explicitly labeled as "fake" or "real," serving as a fundamental binary classification source.

Combined Dataset

Merging these sources **creates a robust dataset of over 50,000 samples,** addressing schema mismatches and class imbalance to ensure high-quality training data.

Data Pipeline: From Raw to Refined

Effective preprocessing is crucial for transforming raw text and metadata into features suitable for machine learning and deep learning models.

System Architecture

Our InfoTrust framework processes news content through a multi-stage pipeline to deliver a comprehensive credibility assessment.

Stage 1: Establishing Baseline with Machine Learning

In the initial phase, we establish performance baselines using classical Machine Learning (ML) models, providing a crucial reference point for the advanced models to be developed in Stage 2.

Model Selection

We employ a suite of classical **ML classifiers: Logistic Regression** (a linear baseline), **Naïve Bayes** (a probabilistic model), **Random Forest** (a tree-based ensemble known for robustness), and **XGBoost** (a powerful gradient boosting algorithm).

Training & Evaluation

Models are trained on the **preprocessed, combined dataset**. Performance is evaluated using standard metrics: **Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC). Hyperparameter tuning, such as GridSearchCV,** is applied to optimize model performance and mitigate overfitting.

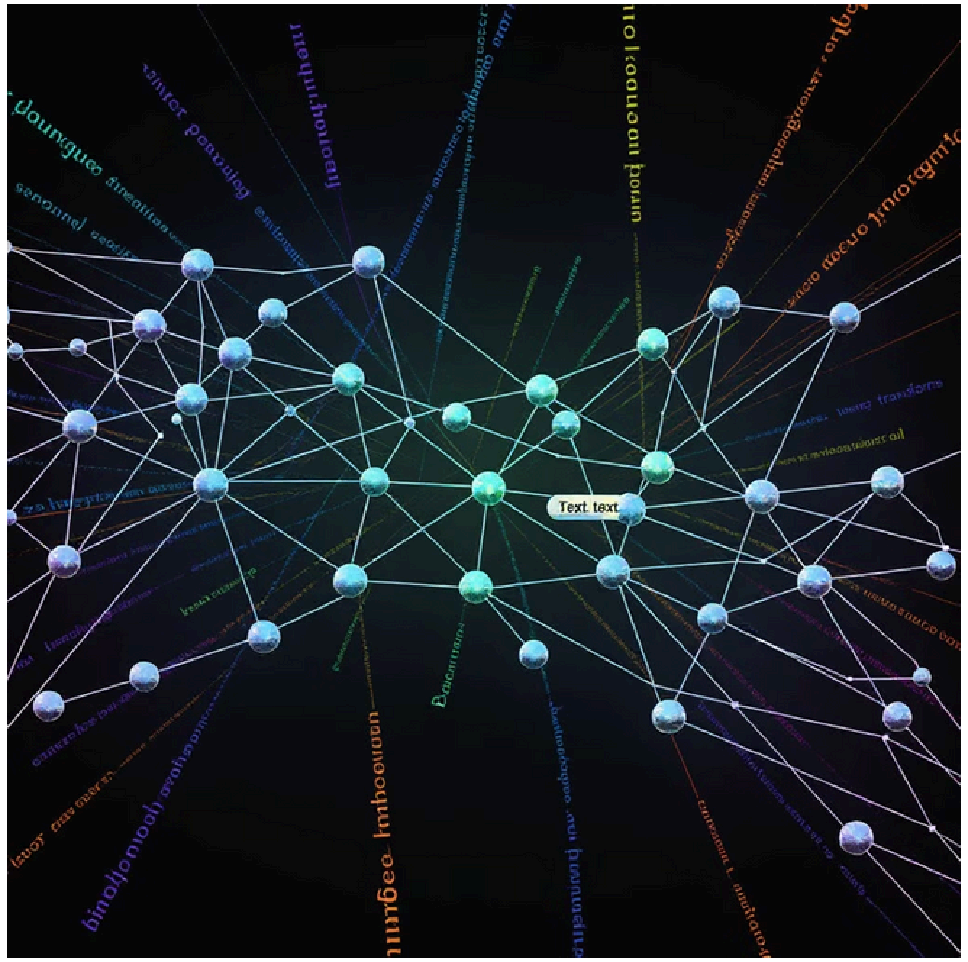
Performance Benchmarking

The results from these ML models provide a baseline performance comparison. This allows us to **quantify the improvements achieved by more complex Deep Learning and Transformer models in subsequent stages.**

These models serve as a foundational step, helping us understand the inherent challenges and initial performance ceilings before moving to more complex architectures.

Stage 2: Advancing with Deep Learning & Transformers

Building on our baseline, **Stage 2 introduces advanced Deep Learning (DL) and Transformer models to capture more complex textual patterns and contextual information** for enhanced credibility analysis.



Deep Learning Models

We implement **BiLSTM (**Bidirectional Long Short-Term Memory) and **GRU** (Gated Recurrent Unit) networks. These models excel at capturing long-term dependencies and sequential information in text, overcoming the vanishing gradient problem common in traditional RNNs.

Transformer Models

We fine-tune state-of-the-art Transformer architectures such as **BERT** (Bidirectional Encoder Representations from Transformers) and **RoBERT**a (Robustly optimized BERT approach). These models leverage attention mechanisms to understand contextual relationships between words, leading to highly accurate classifications.

Multimodal Fusion

A key advancement is exploring multimodal fusion**, combining insights from text, metadata (like user interactions),** and potentially images to create a richer feature set for our models.

These advanced models are crucial for achieving **state-of-the-art performance and overcoming the limitations of simpler ML approaches, especially in capturing nuanced language patterns indicative of misinformation.**

Beyond Content: Source Credibility Analysis

A crucial extension in Stage 2 is integrating source credibility, moving beyond just the content to evaluate the trustworthiness of news origins and propagation networks.

Domain Credibility Scoring

We **assign credibility scores to news domains and sources** based on their historical reliability, **alignment with fact-checking organizations, and past instances of spreading misinformation.** This is a vital meta-feature for the overall assessment.

User Engagement Credibility

Analyzing user interactions (likes, shares, comments) and their networks provides insights into how misinformation spreads. **Graph Neural Networks (GNNs) like GCN (Graph Convolutional Networks) and GAT (Graph Attention Networks) can model these propagation patterns.**

Author Reliability

Evaluating the past record of authors, including their publication history and any known biases or previous involvement in spreading false information, contributes to a more holistic credibility score.

This holistic approach to credibility analysis helps identify coordinated disinformation campaigns and assess the trustworthiness of the entire news ecosystem, not just individual articles.

Explainable AI (XAI): Trust and Transparency

To build a trustworthy system, especially for sensitive applications like credibility analysis, understanding *why* a model makes a certain prediction is as important as the prediction itself.

1

LIME & SHAP

We **utilize LIME** (Local Interpretable Model-agnostic Explanations) and **SHAP** (SHapley Additive exPlanations) to explain individual predictions. These techniques attribute feature importance to words or phrases, showing their contribution to the model's output.

2

Attention Heatmaps

For **Transformer models like BERT, we generate attention heatmaps.** These visual representations highlight which words or tokens the model "focused" on when making its classification, providing intuitive insights into the model's decision-making process.

3

Model Interpretability

By implementing XAI, our framework transforms from a black-box system into a transparent, understandable tool. This is crucial for research credibility and for users to trust the credibility scores provided.

Future Horizons & Project Impact

The InfoTrust framework has significant potential for further expansion and real-world impact.

Real-time Monitoring

Integrate with social media APIs (e.g., **Twitter, Reddit)** for real-time ingestion and credibility analysis of trending news. This moves beyond batch processing to proactive detection.

Multi-Lingual Support

Extend the framework to support multiple languages using multilingual Transformer models (e.g., mBERT, XLM-R), addressing misinformation globally.

Platform Integration

Develop browser extensions or direct integrations with social media platforms, providing instant credibility assessments to users as they browse news feeds.

Scalable Deployment

Transition the prototype (Streamlit/Django web app) to a robust, cloud-based deployment (AWS, Heroku) for high availability and scalability, supporting large-scale usage.

Societal Impact & Research Contribution

InfoTrust aims to empower journalists, policymakers, and the public with tools to navigate the complex information landscape. **Our multi-dataset, credibility-aware, and explainable framework offers a significant research contribution to the fight against misinformation.**

References

- F. G. Hussain, M. Wasim, S. Hameed, A. Rehman, M. N. Asim, and A. Dengel,“Fake News Detection Landscape: Datasets, Data Modalities, AI Approaches, Their Challenges, and Future Perspectives,”**IEEE Access**, vol. 13, pp. 54757–54770, 2025.DOI: 10.1109/ACCESS.2025.3553909

- A. Shu, S. Sliva, S. Wang, J. Tang, and H. Liu,“Fake News Detection on Social Media: A Data Mining Perspective,”**ACM SIGKDD Explorations Newsletter**, vol. 19, no. 1, pp. 22–36, 2017.DOI: 10.1145/3137597.3137600

- K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu,“Fakenewsnet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media,”**arXiv:1809.01286**, 2018.Link: https://arxiv.org/abs/1809.01286

- Z. Zhou, J. Wu, and W. Zeng,“Fake News Detection via NLP is Vulnerable to Adversarial Attacks,”**Proceedings of ACL**, 2020.Link: https://aclanthology.org/2020.acl-main.436/

THANK YOU