Project Proposal: Fake News Detection Using the LIAR Dataset

Overview

Our project focuses on developing a machine learning-based system for fake news detection, using the LIAR dataset, which contains 12.8K short statements labelled by fact-checkers on Politifact.com. Given the rapid spread of misinformation in the digital age, we aim to create a system that can effectively classify the truthfulness of statements, providing a basis for automated fact-checking.

Data Source

We will use the **LIAR dataset** from Hugging Face, which includes:

- **12,836 labelled statements** assessed by Politifact.com.
- **Features/Columns**:
    - **Text Data**: The "statement" field contains the primary text data.
    - **Label**: Truthfulness level (e.g., “true,” “false,” “half-true”) that serves as the target variable.
    - **Additional Metadata**: Columns like "subject," "speaker," "job title," "party affiliation," and historical truthfulness counts, which provide contextual information.

This dataset is well-suited for our project, as it includes both the statement text and contextual features that could potentially enhance classification performance.

Pre-processing Methods

We will implement the following pre-processing steps:

1. **Handle Missing Values**: Identify any missing values in critical fields (e.g., "job title" or "party affiliation") by either filling them in or removing incomplete rows.
2. **Text Cleaning**: For the "statement" column, clean the text by: Removing special characters, Converting text to lowercase, Tokenizing sentences for structured input.
3. **Label Encoding**: Convert categorical truthfulness labels into numerical values suitable for machine learning algorithms.
4. **Feature Engineering**: Include metadata columns like "subject," "party affiliation," and "justification" as additional features to capture more context.
5. **Balancing the Dataset**: Analyze label distribution and, if necessary, apply sampling techniques to ensure balanced classes for model training.

Machine Learning-Based Methods

Our approach will include both traditional and advanced machine learning models:

1. **Baseline Models**:
    - **Logistic Regression**: A basic classifier to establish a benchmark for comparison.
    - **Support Vector Machines (SVM)**: Known for effectively handling high-dimensional data.

    - **Random Forests**: An ensemble method that can capture non-linear relationships and is robust to overfitting.
2. **Advanced Models**:
    - **Transformer Models (BERT)**: Pre-trained language models such as BERT, fine-tuned for our task, will likely excel at capturing nuanced linguistic information in the "statement" text.
    - **RNN/LSTM Models**: RNN-based models will allow us to capture the sequential nature of text, which could be useful for understanding complex language patterns associated with truthfulness or deception.

We aim to compare the performance of these models and examine how much improvement advanced models can provide over traditional methods.

Performance Metrics

To evaluate our models, we will use the following metrics:

- **Accuracy**: Measures the percentage of correctly classified statements.
- **Precision**: Evaluates the proportion of true positives among all predicted positives, indicating the model's accuracy in predicting true labels.
- **Recall**: Assesses the model's ability to identify all actual positives in the dataset, capturing its effectiveness in detecting true labels.
- **F1-Score**: Combines precision and recall, providing a balanced metric that is especially useful when dealing with imbalanced classes.

Related Work and Novelty

The field of fake news detection has seen numerous studies focusing on machine learning methods, as exemplified in the following works:

- Xu, Cheng & Kechadi, Tahar. (2024). An Enhanced Fake News Detection System With Fuzzy Deep Learning. IEEE Access. 12. 88006 - 88021. https://dx.doi.org/10.1109/ACCESS.2024.3418340
- Ning, Xuefei & Wang, Zifu & Li, Shiyao & Lin, Zinan & Yao, Peiran & Fu, Tianyu & Blaschko, Matthew & Dai, Guohao & Yang, Huazhong & Wang, Yu. (2024). Can LLMs Learn by Teaching? A Preliminary Study. https://dx.doi.org/10.48550/arXiv.2406.14629
- Yuchen Zhang, Xiaoxiao Ma, Jia Wu, Jian Yang, and Hao Fan. 2024. Heterogeneous Subgraph Transformer for Fake News Detection. In Proceedings of the ACM Web Conference 2024 (WWW '24). Association for Computing Machinery, New York, NY, USA, 1272–1282. https://doi.org/10.1145/3589334.3645680

Our approach is novel in its combination of both traditional models (for baseline performance) and deep learning models (for advanced, nuanced language understanding) applied specifically to the LIAR dataset. By testing a variety of models and incorporating metadata-based feature engineering, we hope to develop a system that contributes meaningfully to the field of automated fact-checking, with potential applications in social media monitoring and online content moderation.