Objective

The primary objective of this project is to develop and evaluate a neural network model capable of accurately distinguishing between fake and real news articles. By leveraging natural language processing (NLP) techniques and deep learning architectures, the goal is to create a robust classifier that can assist in mitigating the spread of misinformation.

Problem Statement

The proliferation of fake news has significant negative impacts on society, including the erosion of public trust and the skewing of democratic processes. Traditional methods of fact-checking are often labor-intensive and time-consuming. Therefore, there is a critical need for automated systems that can efficiently and effectively identify potentially false information.

Methodology

Data Collection and Preprocessing

Two datasets were used in this project:

1. WELFake Dataset: Contains textual content labeled as real or fake.
2. Fake and Real News Dataset: Comprises separate files for fake and real news articles.

The preprocessing steps involved:

- Concatenation of Title and Text: For some datasets where separate fields for titles and text existed, these were concatenated to form a single text field, ensuring comprehensive feature extraction.
- Cleaning and Normalization:
    - Removing special characters and numbers.
    - Converting all text to lowercase to maintain consistency.
    - Removing stopwords to reduce noise and focus on meaningful words.
    - Lemmatization to reduce words to their base or dictionary form.

Text Vectorization

We used the GloVe (Global Vectors for Word Representation) model for embedding words into numerical vectors. The GloVe model, pre-trained on a large corpus, maps words into a high-dimensional space where the distance and direction between vectors capture semantic relationships between words.

Model Architecture

The model architecture is a hybrid neural network consisting of:

- Embedding Layer: Maps each word to a 300-dimensional vector, using the pre-trained GloVe embeddings.
- Spatial Dropout: Reduces overfitting by dropping entire 1D feature maps in the embedding space, enhancing model generalization.
- Convolutional Layer: Extracts higher-level features through filters that process parts of the input text.
- Bidirectional LSTM: Processes the text in both forward and reverse directions, capturing dependencies from both past and future contexts.
- Global Average Pooling: Reduces the dimensionality of the feature map while retaining important information.
- Dense Layers and Batch Normalization: A fully connected layer with batch normalization follows, introducing non-linearity and scaling the inputs to stabilize and speed up training.
- Output Layer: Comprises two units with a sigmoid activation function, providing the probabilities for the two classes (fake and real).

Training

The model was compiled with the RMSprop optimizer and binary cross-entropy loss function, suitable for binary classification tasks. Training involved several epochs where the model learned to minimize the loss function, adjusting weights through backpropagation.

Results

The model achieved the following performance metrics:

- Training Accuracy: Approximately 93.18%

- Validation Accuracy: Approximately 90.58%
- Test Accuracy: Approximately 90.33%

These results suggest that the model is fairly accurate and generalizes well over unseen data from the test set.

Evaluation and Testing

Additional testing on a concatenated dataset showed a lower accuracy (about 59.90%), indicating potential overfitting or discrepancies in the way the training and external testing datasets were processed or curated.

References

1. Word Embeddings:
    - Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation.
2. Model Architecture:
    - Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory.
    - Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification.
3. Datasets:
    - “WELFake Dataset,” Kaggle.
    - “Fake and Real News Dataset,” Kaggle.

Conclusion

The developed model demonstrates a high degree of accuracy in classifying news articles as fake or real. However, further improvements can be explored by experimenting with different architectures, tuning hyperparameters, or using a more diverse and extensive dataset. Future work should also consider the integration of this model into real-world applications, where it can provide immediate benefits in detecting and flagging fake news content.