Fake News Detection

Prajusha Reddy, Boston University

GOAL

The goal of this project is to be able to classify whether a given article is fake or real news, and also to determine which type of Natural Language Processing (NLP) model performs the best for this classification task. By analyzing the text, the models in this project should be able to determine the credibility of the information. I chose this as my project because I believe that it is more important now than ever before to be able to distinguish between fake and real news; in today's digital age, misinformation spreads rapidly through the use of the internet and different social media platforms (I have seen this happen first-hand through conversations with family, friends, peers, etc.), which is why it is essential to develop tools to identify accurate information for people to make fully informed and smart decisions.

METHODS

Models:

1. Naïve Bayes Model (Baseline)
2. Deep Learning Models
    a. Long Short-Term Memory (LSTM) Model
    b. Bidirectional Encoder Representations from Transformers (BERT) for Sequence Classification Model

RESULTS

```
(base) prajusha@crc-dot1x-nat-10-239-39-106 505project % python train_lstm.py
Epoch 1/20 - Train Loss: 0.6803 | Val Loss: 0.6653
Epoch 2/20 - Train Loss: 0.6622 | Val Loss: 0.6114
Epoch 3/20 - Train Loss: 0.4161 | Val Loss: 0.3676
Epoch 4/20 - Train Loss: 0.3654 | Val Loss: 0.3070
Epoch 5/20 - Train Loss: 0.2781 | Val Loss: 0.2212
Epoch 6/20 - Train Loss: 0.2667 | Val Loss: 0.2337
Epoch 7/20 - Train Loss: 0.2164 | Val Loss: 0.2835
Epoch 8/20 - Train Loss: 0.2247 | Val Loss: 0.1805
Epoch 9/20 - Train Loss: 0.1713 | Val Loss: 0.1462
Epoch 10/20 - Train Loss: 0.1755 | Val Loss: 0.2040
Epoch 11/20 - Train Loss: 0.1935 | Val Loss: 0.1862
Accuracy: 0.9635
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96      3753
           1       0.97      0.96      0.97      4364

    accuracy                           0.96      8117
   macro avg       0.96      0.96      0.96      8117
weighted avg       0.96      0.96      0.96      8117
```

LSTM Model Evaluation Metrics

```
Test Set Classification Report:
              precision    recall  f1-score   support

           0     0.9890    0.9803    0.9846      3753
           1     0.9832    0.9906    0.9869      4364

    accuracy                         0.9858      8117
   macro avg     0.9861    0.9854    0.9857      8117
weighted avg     0.9859    0.9858    0.9858      8117
```

BERT Model Evaluation Metrics

```
Accuracy: 0.9390
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.90      0.93      3753
           1       0.92      0.97      0.94      4364

    accuracy                           0.94      8117
   macro avg       0.94      0.94      0.94      8117
weighted avg       0.94      0.94      0.94      8117
```

Naïve Bayes Model Evaluation Metrics

EXPERIMENTS

Naïve Bayes Model

- Accuracy: 93.9%

LSTM Model

- Accuracy: 96.35%

BERTForSequenceClassification Model

- Accuracy: 98.58%

As expected, the Naïve Bayes has the lowest accuracy and overall evaluation metrics compared to the LSTM and BERT models, despite having decent performance.

CONCLUSIONS

Through my experiments, I learned that while a simple model like Naïve Bayes can achieve decent results with minimal computation, deep learning models, such as the LSTM and BERT models, significantly outperform it in regards to text classification, such as fake news classification. The BERT model especially, being a Transformer model, yields very good results. Additionally, I also observed that the BERT model achieved very high accuracy with very few training epochs, and it may have even better results with just one epoch, which showcases its dominance in text classification tasks.