Fake News Detection Using BERT

Group 16: Jing Xia, Siyuan Liu, Yang Zhao

Table of contents

01

Introduction

Motivation

Why fake news detection is important?

- Preventing the Spread of Misinformation
- Protecting Public Trust
- Protecting Consumers and Society

Goal

- Design and train a classification model to identify fake news articles using BERT
- Analyze and evaluate the performance of the model
- Explore the impact of different configurations on model performance using ablation study

02

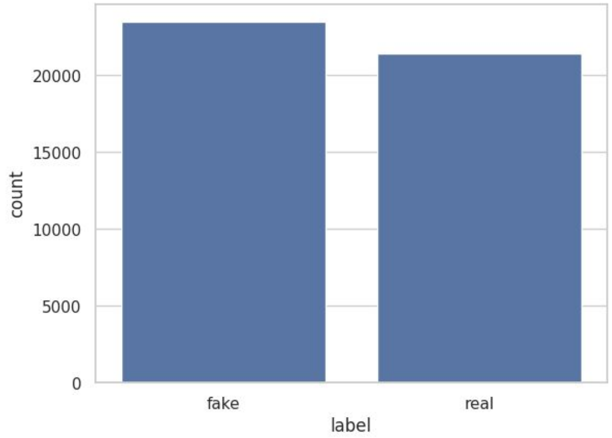
Dataset

Source and structure

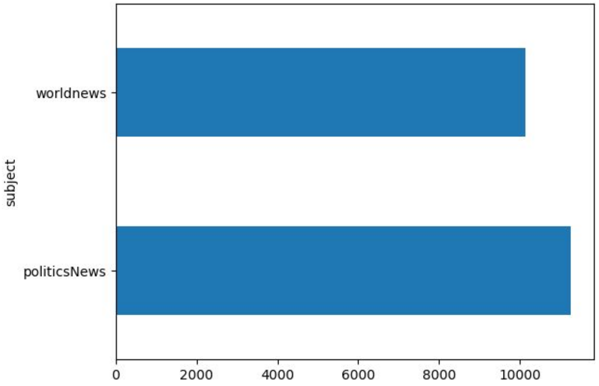
- Kaggle's Fake News Detection Dataset
- Fake.csv: Contains fabricated news articles
- True.csv: Contains authentic news articles from credible sources

Exploratory Data Analysis (EDA)

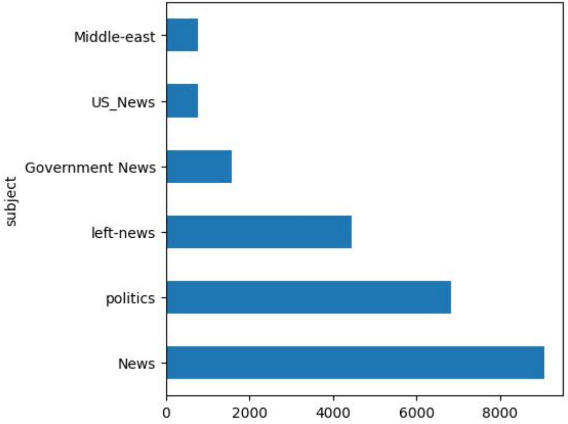
Class Distribution



Subject Distribution



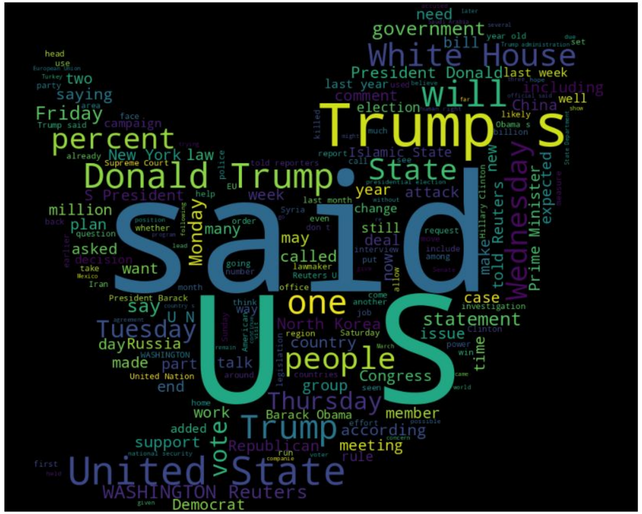
Real news



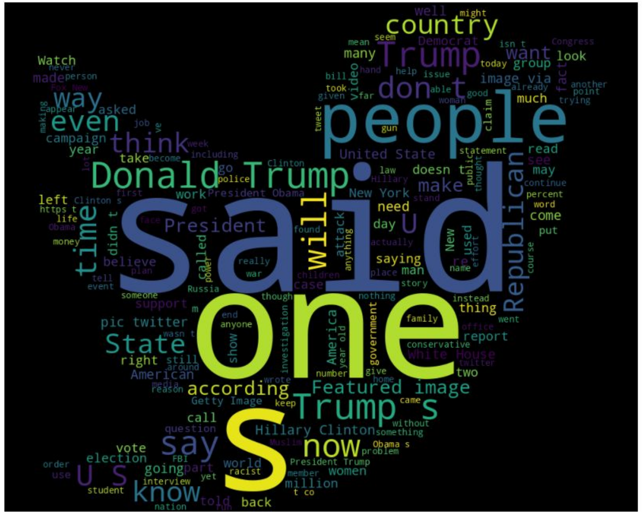
Fake news

Exploratory Data Analysis (EDA)

Word Clouds



Real news



Fake news

Data Preprocessing

- **Data Cleaning**: Drop unnecessary columns, combine text and title columns
- **Tokenization**: Use BERT tokenizer to convert text into a sequence of tokens preserving the contextual meaning of words
- **Handling Imbalanced Data**: Sample 1000 real news and 1000 fake news

03

Model Architecture

Architecture

Layer (type)	Output Shape	Param #	Connected to
attention_mask (InputLayer)	(None, 100)	0	-
input_ids (InputLayer)	(None, 100)	0	-
bert_layer (BertLayer)	(None, 768)	0	attention_mask[0][0], input_ids[0][0]
dropout (Dropout)	(None, 768)	0	bert_layer[0][0]
dense (Dense)	(None, 64)	49,216	dropout[0][0]
dropout_1 (Dropout)	(None, 64)	0	dense[0][0]
dense_1 (Dense)	(None, 1)	65	dropout_1[0][0]

```
Total params: 49,281 (192.50 KB)
Trainable params: 49,281 (192.50 KB)
Non-trainable params: 0 (0.00 B)
```

Training and Hyperparameter Tuning

Batch size: Experimented with values of 16, 32, and 64.

Learning rate: Varied from 1e-04, 1e-05 and 1e-06 to optimize convergence.

Optimizers: Adam, SGD, Nadam.

27 Combinations

Try different combinations for parameters

9 parameters

Including batch size, learning rate, and optimizer

Evaluation Metrics

Feature	Description
Accuracy	Percentage of correctly classified articles.
Precision	Fraction of true positives among predicted positives.
Recall	Fraction of true positives among actual positives.
F1-Score	Harmonic mean of precision and recall.
Confusion Matrix	Visualized true positives, false positives, true negatives, and false negatives.

04

Results

70.2%

Accuracy

70%

Precision

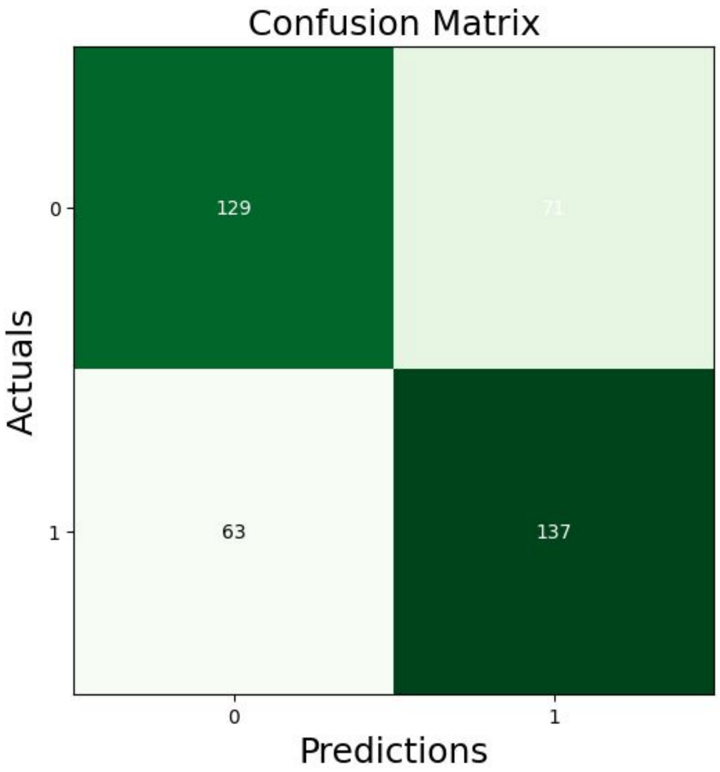
71%

Recall

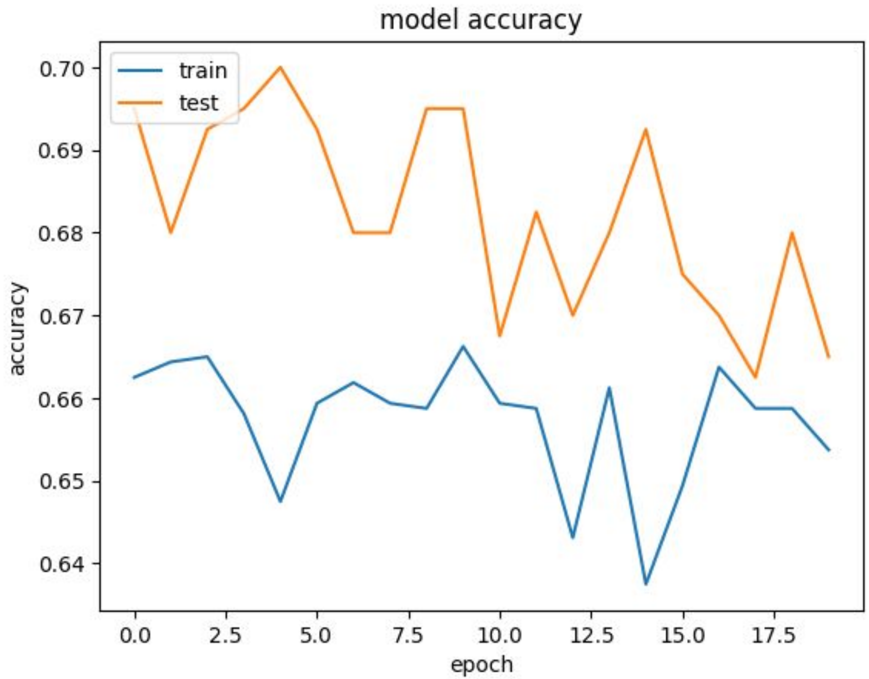
73%

Under the curve area

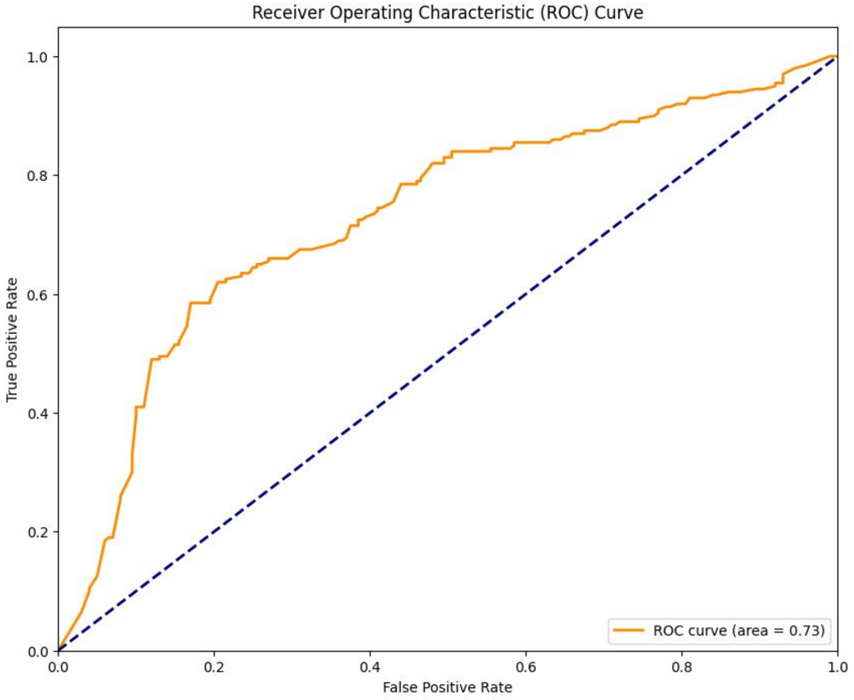
Confusion Matrix



Model Accuracy vs Epochs



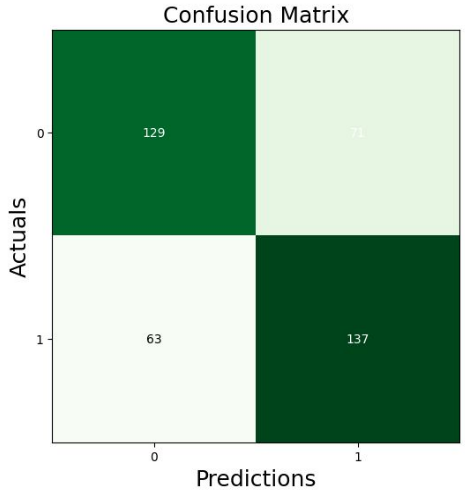
ROC Curve with AUC = 0.73



05

Analysis

Analysis



Strengths:

- High precision in identifying real news.
- Effective handling of long-form articles,

Weakness:

- Lower recall on fake news with subtle writing styles.
- Limited generalization to datasets outside the training domain.

Summary:

Overall, the model provides a strong foundation for future enhancements and specific use case applications.

06

Ablation Study

Ablation study

Text Preprocessing Methods

- Stop words removal
- Lemmatization
- Stemming

	precision	recall	f1-score	support
0	0.71	0.64	0.67	200
1	0.67	0.73	0.70	200
accuracy			0.69	400
macro avg	0.69	0.69	0.69	400
weighted avg	0.69	0.69	0.69	400

Stop words

	precision	recall	f1-score	support
0	0.70	0.79	0.74	200
1	0.76	0.66	0.71	200
accuracy			0.72	400
macro avg	0.73	0.73	0.72	400
weighted avg	0.73	0.72	0.72	400

Lemmatization

	precision	recall	f1-score	support
0	0.56	0.41	0.48	200
1	0.54	0.68	0.60	200
accuracy			0.55	400
macro avg	0.55	0.55	0.54	400
weighted avg	0.55	0.55	0.54	400

Stemming

Ablation study

BERT Tokenization and Max Length

- Max length of 100
- Max length of 128

	precision	recall	f1-score	support
0	0.67	0.65	0.66	200
1	0.66	0.69	0.67	200
accuracy			0.67	400
macro avg	0.67	0.67	0.66	400
weighted avg	0.67	0.67	0.66	400

Max length of 100

	precision	recall	f1-score	support
0	0.65	0.54	0.59	200
1	0.61	0.71	0.66	200
accuracy			0.62	400
macro avg	0.63	0.62	0.62	400
weighted avg	0.63	0.62	0.62	400

Max length of 128

Ablation study

Activation Functions in the Classification Layer

- Using Sigmoid
- Using Softmax

	precision	recall	f1-score	support
0	0.62	0.69	0.65	200
1	0.65	0.58	0.61	200
accuracy			0.64	400
macro avg	0.64	0.64	0.63	400
weighted avg	0.64	0.64	0.63	400

Using Sigmoid

	precision	recall	f1-score	support
0	0.00	0.00	0.00	200
1	0.50	1.00	0.67	200
accuracy			0.50	400
macro avg	0.25	0.50	0.33	400
weighted avg	0.25	0.50	0.33	400

Using Softmax

07

Conclusion

Conclusion

- BERT has proven to be a powerful model for text classification tasks like fake news detection.
- Its ability to capture contextual embeddings enables high accuracy and precision, making it a valuable tool in NLP applications.

Future work

- **Data Augmentation**: Introduce synthetic examples to improve generalization.
- **Ensemble Models**: Combine BERT with other NLP models to enhance robustness.
- **Explainability**: Integrate techniques like SHAP to interpret model predictions and uncover bias.

Thank you!