

Fake News Detection Using Machine Learning

Using BERT and TF-IDF + Logistic Regression

Name: Collin Kerker

Problem, Motivation, and Objective

Problem: Fake news spreads very quickly online and can cause a lot of damage

- Anyone can publish information in any way without fact checking

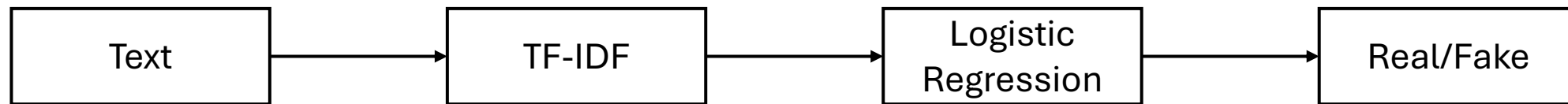
Objective: Build a model that can detect whether a news article is fake or real

Dataset Overview

- Dataset: “Fake and Real News Dataset” (Kaggle)
- ~40,000 labeled articles total
- Columns: title, text, subject, date
- Train/test split is 80/20

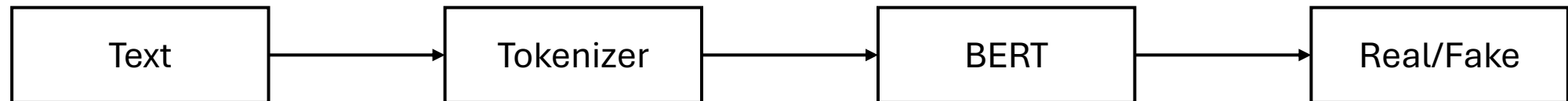
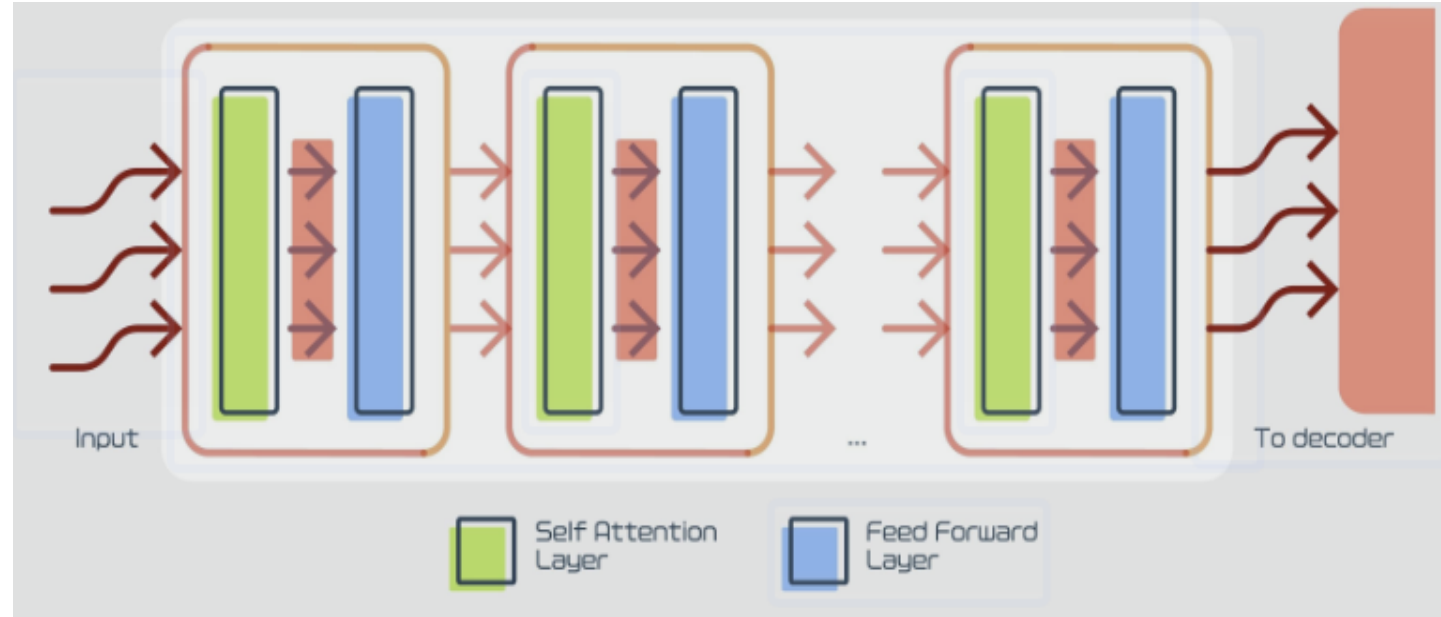
Baseline Model: TF-IDF + Logistic Regression

- TF-IDF converts text into numerical features based on word importance
- Logistic Regression works fast and works well with high-dimensional text features, and is a staple of text classification
- This is a strong baseline for fake vs real news detection



BERT Model: bert-base-uncased

- This model is pretrained on large English texts, like Wikipedia
- Understands context from the words
- Takes in tokenized text with a max length of 256 tokens
- Outputs either 0 or 1 for fake or real



Training Setup

- HuggingFace Transformers (handles tokenizer, the model, the datasets, and the training loop through trainer API) + PyTorch
- Trained on CPU
- Parameters:
 - 3 epochs – Anymore would take too long on a CPU
 - Batch size of 8 – Does not overload CPU memory with BERT
 - Learning rate of $2e-5$ – Recommended learning rate for BERT

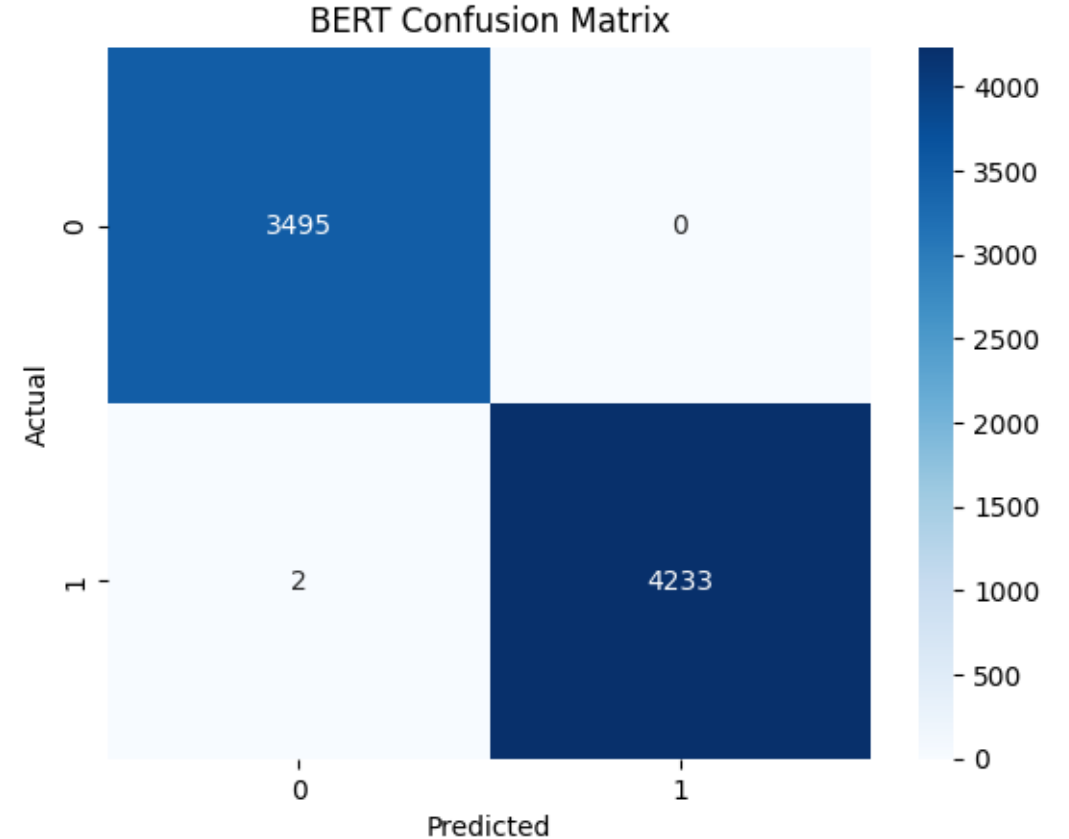
Results: Baseline vs BERT

- Baseline (TF-IDF + Logistic Regression):
 - Accuracy = 0.986, F1 = 0.987
- BERT (bert-base-uncased):
 - Accuracy = 0.9997, F1 = 0.9997
- BERT almost perfectly classified fake vs real articles

	Model	Accuracy	F1 Score
0	Baseline (TF-IDF + Logistic Regression)	0.986417	0.987680
1	BERT	0.999741	0.999764

Confusion Matrix (BERT)

- Fake: 3495 correctly predicted as fake, 0 predicted as real
- Real: 4233 correctly predicted as real, 2 predicted as fake
- Shows BERT is almost perfect on this test set



Why So Good?

- TF-IDF works well because it can latch onto prominent words in the articles, which are used a lot specifically in fake news articles
- TF-IDF can capture strong patterns because of the length of the articles
- BERT is pretrained and therefore captures deep context and word relationships

Limitations and Future work

- Trained on CPU, so limited epochs and no heavy tuning
- BERT is too good
- Speculation: This dataset is likely simpler than real news articles
- For the future:
 - Have more difficult to interpret data
 - Use GPU instead of CPU
 - Test other transformer models

Step	Training Loss
500	0.013000
1000	0.002300
1500	0.008700
2000	0.006900
2500	0.005000
3000	0.004500
3500	0.002800
4000	0.009300
4500	0.005600
5000	0.004400
5500	0.007100
6000	0.000000
6500	0.000000
7000	0.000000
7500	0.000000
8000	0.000000
8500	0.000000
9000	0.000000
9500	0.000000
10000	0.000000
10500	0.000800
11000	0.000000
11500	0.000000

Conclusion

- Built and compared a baseline model and a trained BERT model for fake news detection
- Both performed almost perfectly
- Shows that this was a good task for a TF-IDF + Logistic Regression Baseline
- Shows that this was too easy for BERT