Fact Verification of Politician Statement Comparing Baseline and Fine-Tuned Models

Vaishnav Mandlik
University of Arizona
vaishnavm@arizona.eud

Abstract

Misinformation in politics is a growing problem in today's digital world. To tackle this, I worked on automating fact-checking using machine learning. I used a dataset of political statements labeled as true or false and tested two approaches: a basic logistic regression model and a fine-tuned BERT model. The logistic regression model performed reasonably well, with an F1-score of 66, but the BERT model did better, achieving an F1-score of 71. This paper explains how I prepared the data, the methods I used, my results, and the challenges I encountered.

1 Introduction

The internet makes it easy to share information, but it also makes it just as easy to spread misinformation. In politics, this can seriously affect public opinions and decisions. Fact-checking can help fight this problem, but doing it manually takes a lot of time and effort. That's where machine learning comes in—it can automate fact-checking and make it faster and more efficient.

In this project, I worked on building and comparing two models to classify political statements as true or false. I started with a simple logistic regression model, which is often a good starting point for language tasks. Then, I tested BERT, a more advanced model designed to understand the meaning and context of words. The goal was to see how well these models could handle the tricky language used in politics and understand the challenges of automating fact-checking.

2 Related Work

Fact-checking systems have come a long way over time. Early methods mostly relied on rule-based systems and lexical matching, which had trouble with more nuanced or ambiguous language. With the rise of machine learning, models like logistic regression and SVM became popular for text classification, using features such as TF-IDF and word embeddings.

In recent years, deep learning has changed the landscape of NLP. Transformer models like BERT have shown impressive results across many tasks, such as sentiment analysis and question answering. Research has shown that BERT's ability to understand context gives it an edge over traditional models, especially when dealing with complex language. Our work builds on these advancements by applying BERT to political fact-checking and comparing its performance with a simpler baseline model.

3 Methodology

3.1 Dataset

I used a dataset of 2,874 political statements, each labeled to indicate its truthfulness, ranging from "true" to "pants-on-fire." To simplify the task, I grouped the labels into two categories:

- **True (1):** TRUE, mostly-true
- **False (0):** FALSE, mostly-false, half-true, pants-on-fire

making it a straightforward binary classification problem. The data was split into 90% for training and 10% (less training data due to resources) for testing.

3.2 Preprocessing

For the logistic regression model, I cleaned the text by removing punctuation, converting everything to lowercase, and extracting features like unigrams and bigrams using TF-IDF. For BERT, I used tokenization to prepare the text while keeping its original context intact, taking advantage of BERT's ability to handle subword representations.

3.3 Models

3.3.1 Logistic Regression

This model served as a simple, interpretable baseline. It was trained on the TF-IDF features, focusing on unigram and bigram patterns to capture important word relationships.

3.3.2 BERT

BERT was fine-tuned using the HuggingFace library. A classification head was added to output true/ or false predictions. The model is trained over three epochs.

3.4 Evaluation

To evaluate the models, I used performance metrics including accuracy, precision, recall, F1-score. These

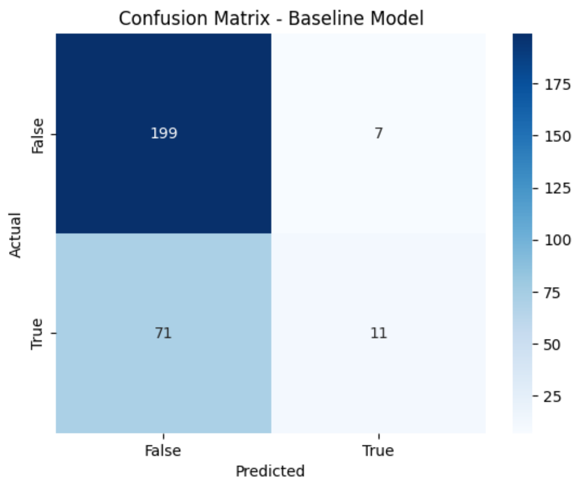

Figure 1: Confusion Matrix of Baseline Model

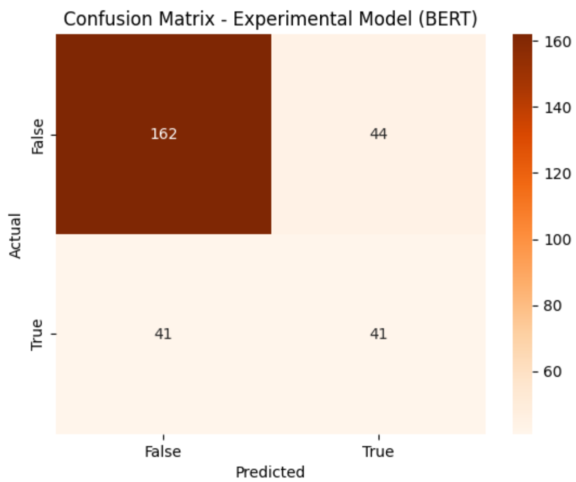

Figure 2: Confusion matrix of Experimental Model

metrics shows the model performance across both true and false labels.

4 Experiments and Results

4.1 Baseline Results

The logistic regression model worked on average, achieving 73% accuracy and an F1-score of 66%. It had unable to identify true statements correctly, with recall for true statements being only 13%. Fig 1 Confusion Matrix

4.2 BERT Results

The fine-tuned BERT model worked well in recall for true statements, reaching 50%. While its accuracy was slightly lower (70%), the BERT model achieved an F1-score of 71%, which was better than the logistic regression model. This shows that BERT was more effective at understanding the politician statements. Fig.2 Confusion matrix

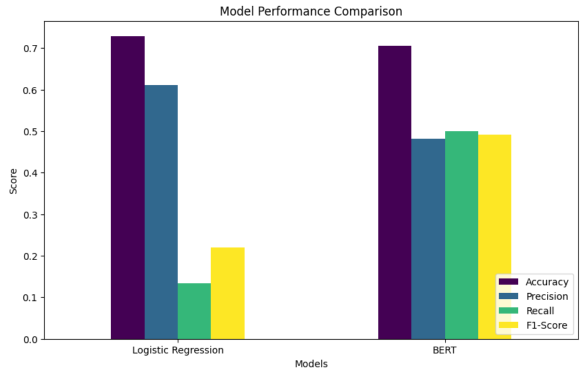

Figure 3: BERT vs Baseline model

5 Experiments and Results

5.1 Comparison

Accuracy Precision, Recall, and F1-Scrore both models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	73%	70%	73%	66%
BERT (fine-tuned)	70%	71%	64%	71%

Table 1: Comparison of model performance

5.2 Model Performance

Both models face difficulties with ambiguous statements, especially those that need external knowledge to interpret. For example, statements like “"Thirty million Americans, including a lot of people in Florida, are going to be able to get healthcare next year because of that law."” like this statement require an understanding of future projections that neither model could fully grasp. fig 3.

6 Conclusion

This project study shows the two machine learning methods for political fact-checking logistic regression and BERT. While logistic regression served as a solid baseline, BERT showed clear strengths in understanding complex language.