A Comparative Analysis of Misinformation Detection in NLP: Traditional Approach vs. RoBERTa

Shashwat Shrivastava

Abstract

This project explores two different approaches for identifying misinformation in news articles: a baseline model built with TF-IDF and Logistic Regression, and a fine-tuned RoBERTa transformer model. Both methods were evaluated for their ability to classify news as real or fake, using metrics like accuracy, precision, recall, F1-score, and confusion matrices. The goal was to compare the performance of a simple, traditional machine learning model with a more advanced transformer-based approach in tackling the challenge of misinformation. Further analysis was also done using a larger dataset to highlight the efficacy of new models.

1 Introduction

In today's digital age, the spread of misinformation has become a significant concern, particularly through online news platforms. Developing automated systems to detect false or misleading content is essential to address this challenge effectively.

This project compares two methods:

- **Baseline Model:** Utilizes TF-IDF to extract features from text and applies Logistic Regression for classification.
- **RoBERTa Model:** A fine-tuned transformer-based model designed specifically for detecting misinformation.

By assessing their performance, the project aims to understand the trade-off between simplicity and advanced capabilities.

2 Objective

The primary objective of this project is to assess and compare the effectiveness of two different models in identifying misinformation in news articles. This involves evaluating their strengths, weaknesses, and overall performance to determine which approach is better suited for the task. The comparison focuses on:

1. Classification accuracy, precision, recall, and F1-score of both models.
2. Comparison of models using confusion matrices and classification reports.
3. Trade-offs between classical machine learning and transformer-based deep learning approaches.

3 Methodology

3.1 Dataset

The original dataset was sourced from Kaggle which has over 43k records of news articles (both real and fake combined). https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets.

For this project, a subset of the dataset was taken which includes 1000 records: 500 real news and 500 fake news articles.

3.2 Data Preparation

The data was cleaned and split into training and testing sets in an 80-20 ratio. The preparation ensured balanced representation for both classes.

3.3 Baseline Model: TF-IDF + Logistic Regression

- • The text data was converted into numerical features using TF-IDF, which measures the importance of words in the dataset.
- • Logistic Regression was used to classify the articles as real or fake based on these features.
- • Performance metrics such as accuracy, precision, recall, and F1-score were calculated to evaluate the model.

3.4 RoBERTa Model

- • The RoBERTa transformer model was fine-tuned using the Hugging Face library to optimize it for the task of detecting misinformation.
- • The model was then tested on the same dataset and evaluated using the same performance metrics as the baseline model.

4 Evaluation

4.1 Baseline Model Results

- Accuracy: 96%
- Precision: 0.96 (both classes)
- Recall: 0.96 (both classes)
- F1-score: 0.96 (both classes)

Baseline Model Performance:

	precision	recall	f1-score	support
0	0.96	0.96	0.96	104
1	0.96	0.96	0.96	96
accuracy			0.96	200
macro avg	0.96	0.96	0.96	200
weighted avg	0.96	0.96	0.96	200

Figure 1: Classification Report for Baseline Model. O refers to fake/misleading news and 1 refers to real/true news

The Logistic Regression model delivered a commendable accuracy of 96% ,offering balanced performance with consistent precision and recall.

4.2 RoBERTa Model Results

- Accuracy: 99%
- Precision: 1.00 (class Fake), 0.99 (class Real)
- Recall: 0.99 (class Fake), 1.00 (class Real)
- F1-score: 1.00 (class Fake), 0.99 (class Real)
- Training Time: 12,020 seconds (approximately 3.34 hours)

RoBERTa Classification Report:

	precision	recall	f1-score	support
0	1.00	0.99	1.00	104
1	0.99	1.00	0.99	96
accuracy			0.99	200
macro avg	0.99	1.00	0.99	200
weighted avg	1.00	0.99	1.00	200

Figure 2: Classification Report for RoBERTa Model.O refers to fake/misleading news and 1 refers to real/true news

5 Analysis

The RoBERTa model demonstrated superior performance, effectively identifying complex patterns and nuances in the data. However, it required significant computational resources and significant extended training time. Figure x and y show the training and testing time required, respectively for the RoBERTa model on the original complete dataset.

In contrast, the baseline TF-IDF with Logistic Regression proved efficient and fast, making it a practical choice for scenarios with limited resources.

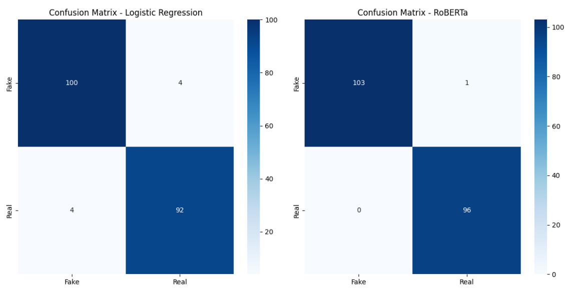


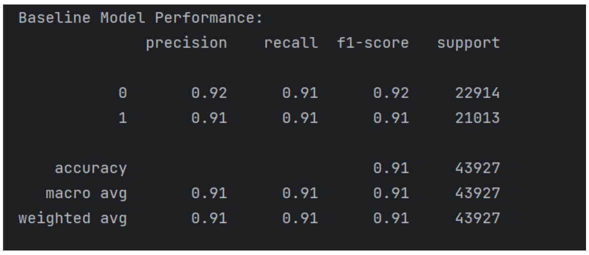
Figure 3: Confusion Matrix for the baseline and RoBERTa model. X- axis represents the actual labels while y axis represents the predicted labels.

5.1 Further Analysis

The original source dataset was used for additional testing to assess the models' performance and scalability. The results are summed up as follows:

1. The results for the RoBERTa model could not be included in this report due to the very restricted computational resources and the extremely high volume of data in the original source dataset.
2. On the larger dataset, the TF-IDF + Logistic Regression model showed 92% accuracy,

demonstrating its effectiveness and capacity to manage large data volumes with less resource usage.


Baseline Model Performance:

	precision	recall	f1-score	support
0	0.92	0.91	0.92	22914
1	0.91	0.91	0.91	21013
accuracy			0.91	43927
macro avg	0.91	0.91	0.91	43927
weighted avg	0.91	0.91	0.91	43927

Figure 4: Classification Report for Baseline model on the entire dataset (Kaggle).O refers to fake/misleading news and 1 refers to real/true news

6 Conclusion

This project highlights the trade-offs between traditional machine learning models and transformer-based models such as RoBERTa to detect misinformation. While RoBERTa excels in accuracy and nuanced understanding, traditional approaches like TF-IDF with Logistic Regression remain valuable for resource-constrained applications. Future work could focus on optimizing RoBERTa's training process to reduce resource consumption or exploring other transformer models to further improve efficiency.

References

1. Hugging Face RoBERTa Documentation. https://huggingface.co/docs/transformers/en/model_doc/roberta

2. Tfidf Vectorizer - Scikit Learn. https://scikit-learn.org/1.5/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

A Appendix

This section contains supplementary figures and materials that support the main content of the paper.

Figure 5: Image showing training time taken (3.34hours) to train the RoBERTa model on the 1000records dataset

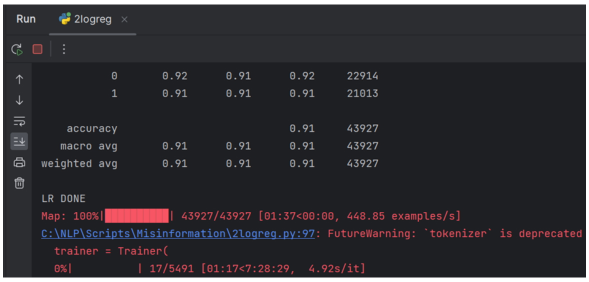


Figure 6: Time required ('7hours) to evaluate the RoBERTa model on the entire original datatset.