DETECTING FAKE NEWS USING MACHINE LEARNING

PRESENTED BY ANNIE MENESES

Introduction

Objective: To build a classifier that determines whether a given headline is fake news or not.

Motivation: The increasing challenge of fake news in today's digital era. Fake news affects public opinion, elections, and even individual lives.

Impact: Headlines are often misleading or fabricated to grab attention, and detecting such content is difficult for humans. An automated system can help reduce the spread of false information and ensure more trustworthy content.

Real word applications:

- Social Media Monitoring. Using machine learning to identify and flag fake news can help improve the integrity of content on these platforms.

- Media Integrity: News organizations can use these systems to check the credibility of articles before publishing or distribute fact-checked content.

- Government: Governments and non-profits can use automated systems to detect and report fake news, especially around sensitive topics like elections or public health crises.

Challenges:

- Language Complexity
- Lack of Context

Approach

Data cleaning and preprocessing:

- Dataset was quite clean with no missing values
- The dataset was quite balanced in fake and not-fake news
- Tokenization, punctuation removing, lemmatization, stop words removing
- Null rows dropping →

After this process headlines with 1 or 2 words turn to null columns, maybe an approach to remove this rows from the beginning could have be implemented

Feature Engineering:

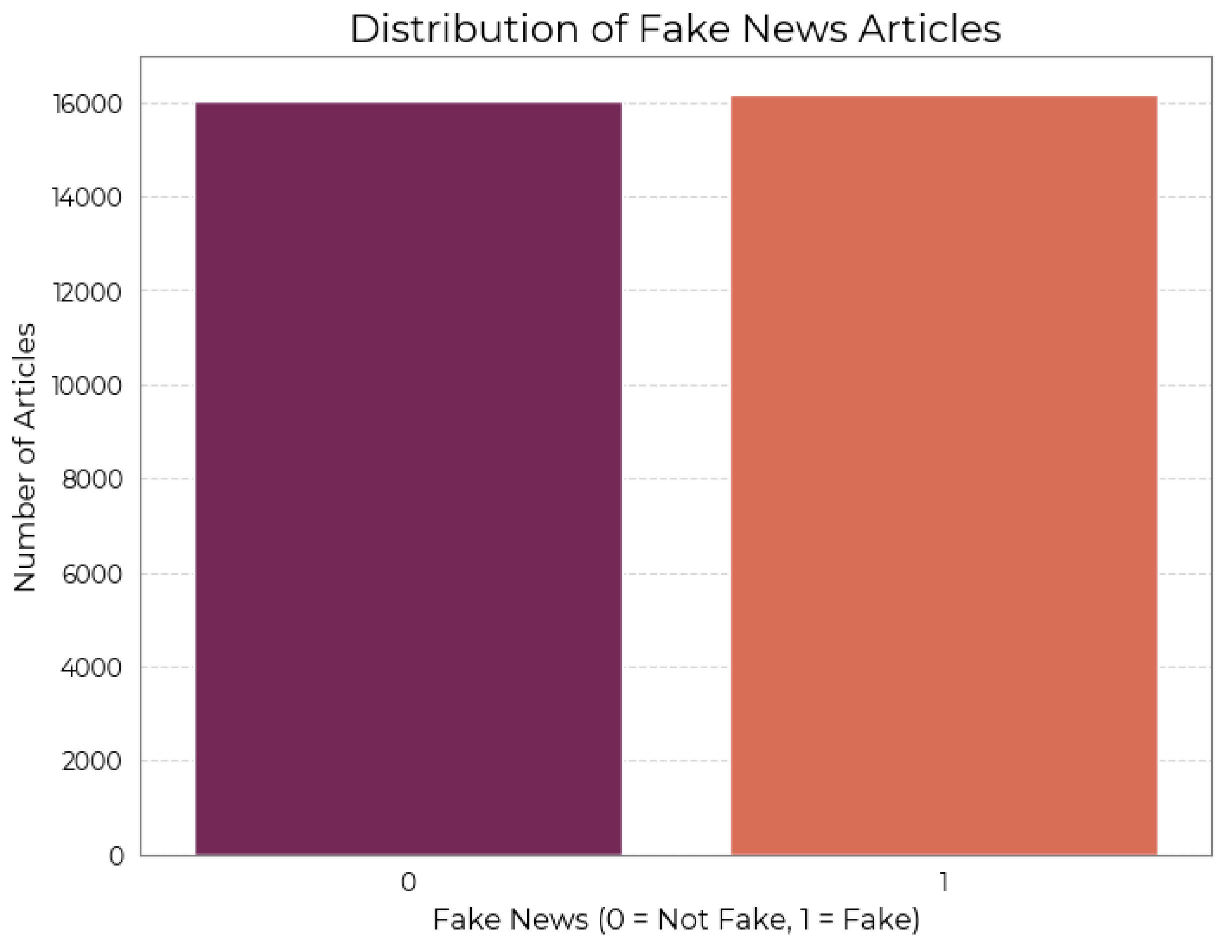
- Sentiment analysis
  - VADER
  - RoBERTa
- Text Vectorization
  - BoW (Bag of Words)
  - Word Embeddings (Word2Vec)
  - Transformer-based
    - BERT → 'bert-base-uncased'
    - MiniLM -> 'all-MiniLM-L6-v2'

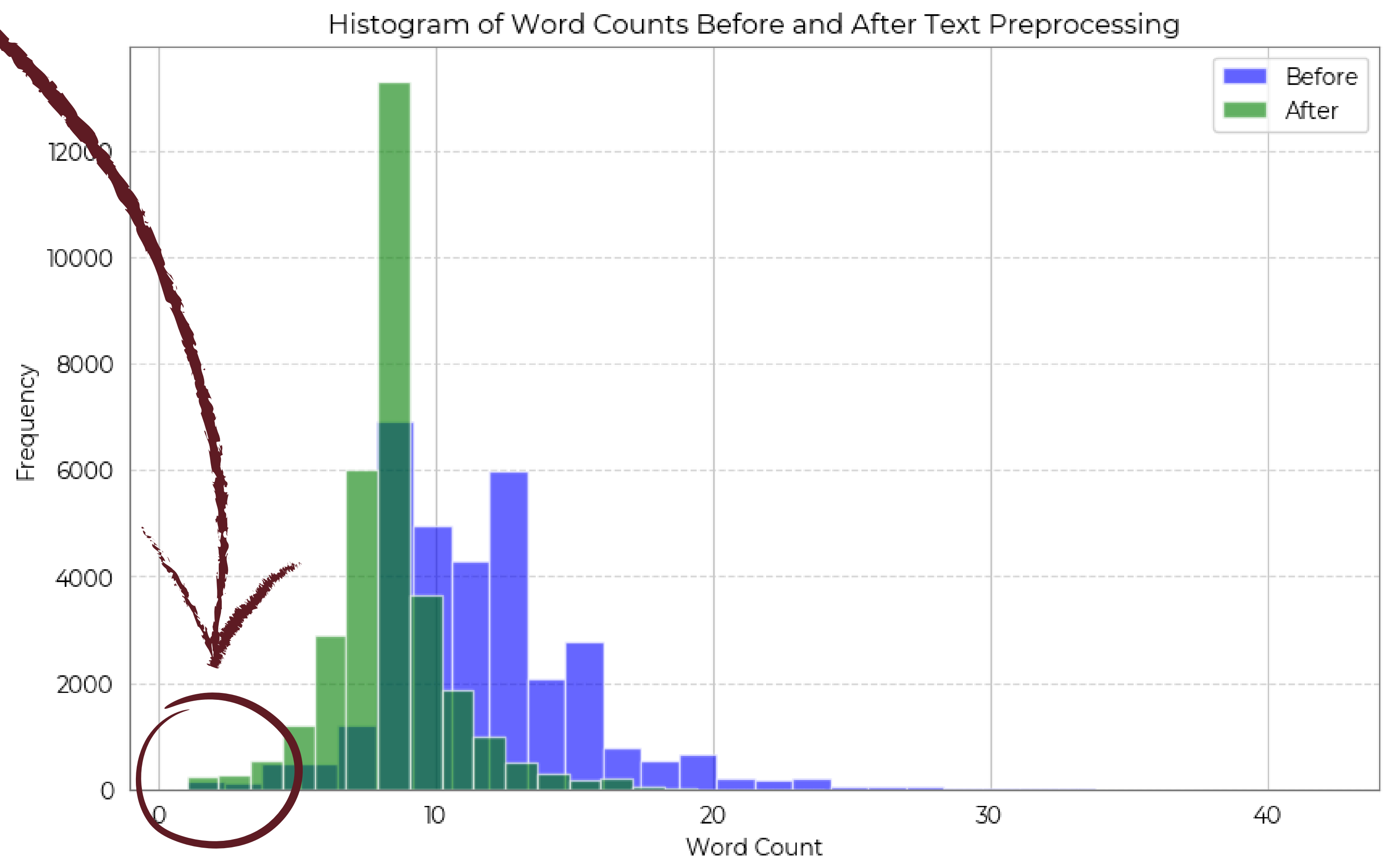
To the headlines before preprocessing

*Would be interesting to check the results in the preprocessed texts

Model Training & Evaluation:

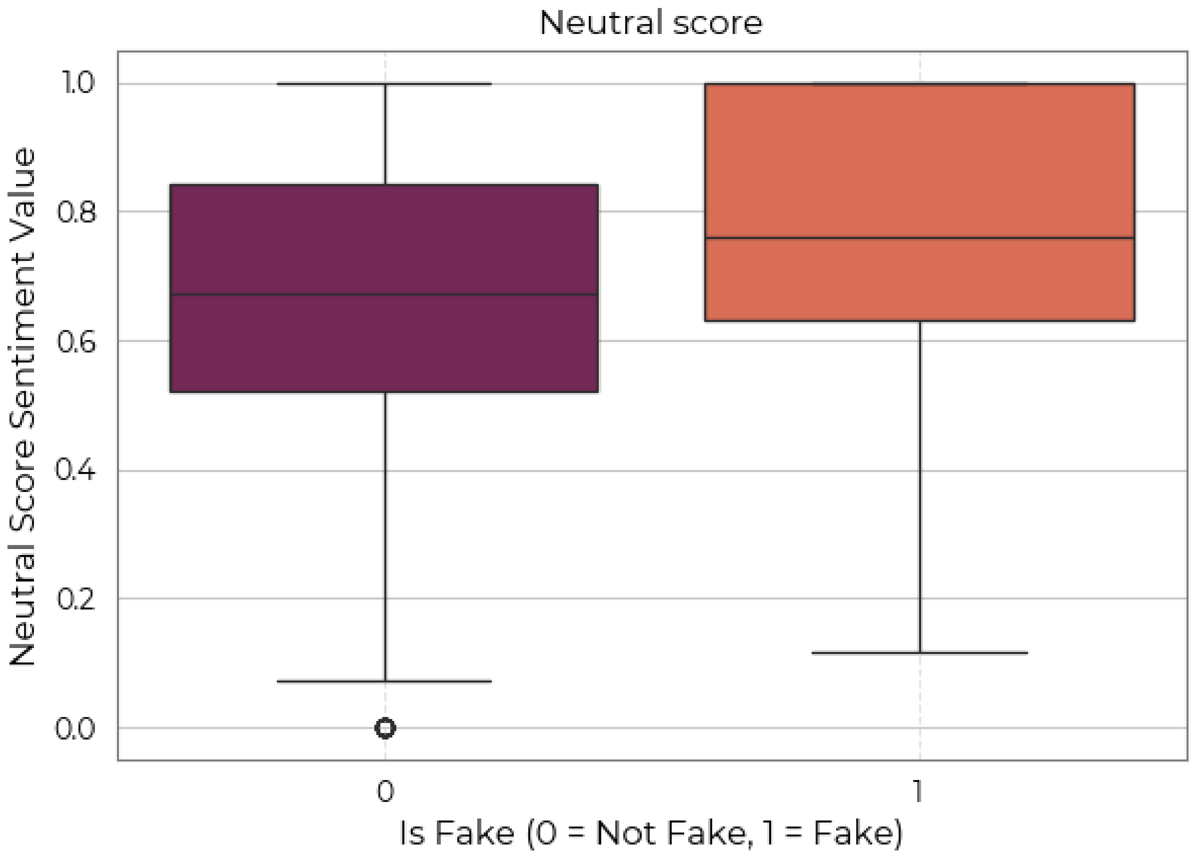
- Logistic Regression
- Random Forest
- SVM

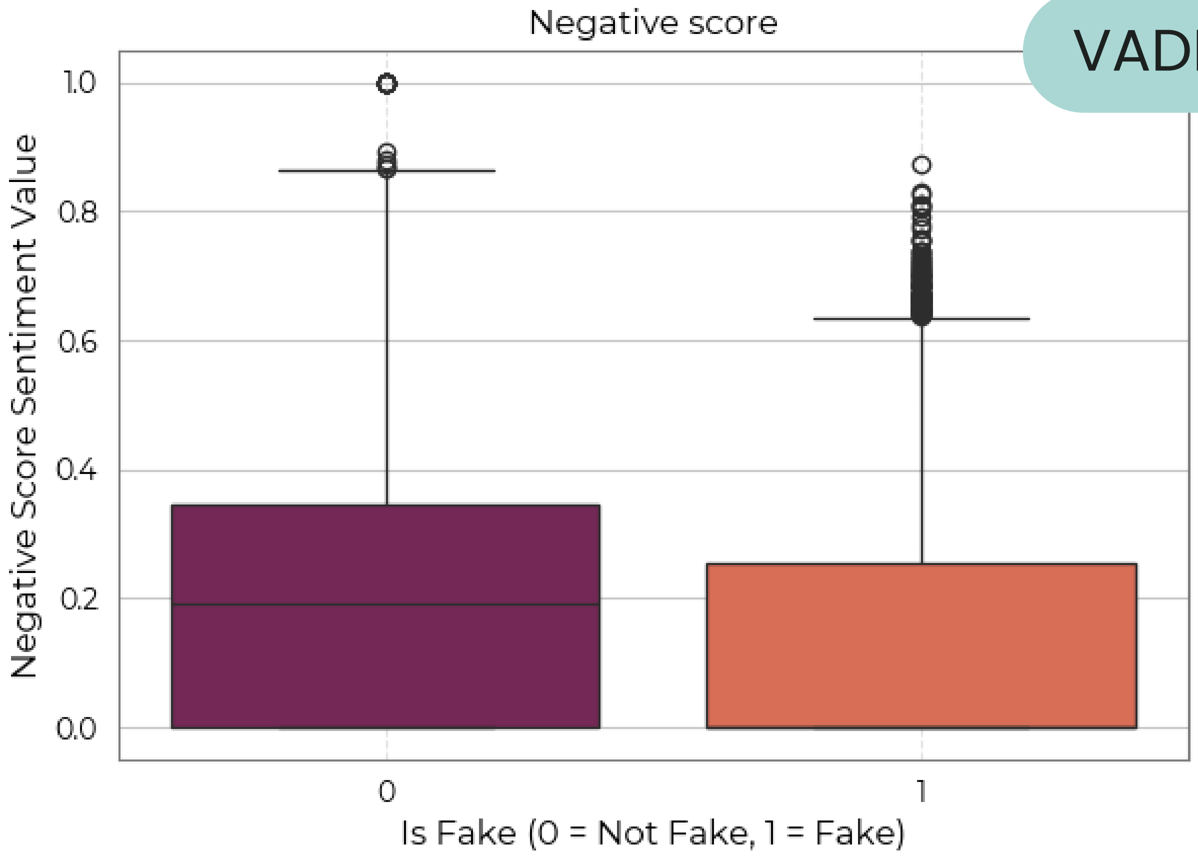




Sentiment analysis

VADER

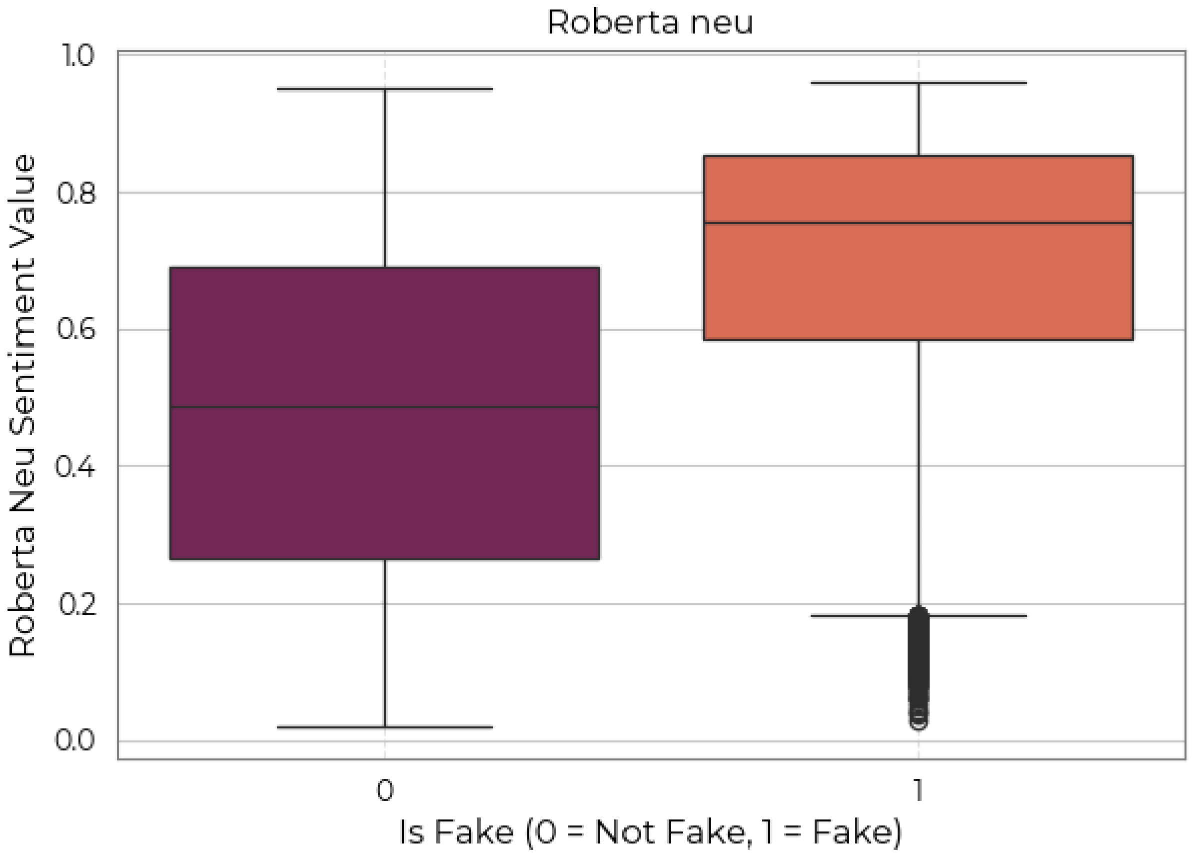


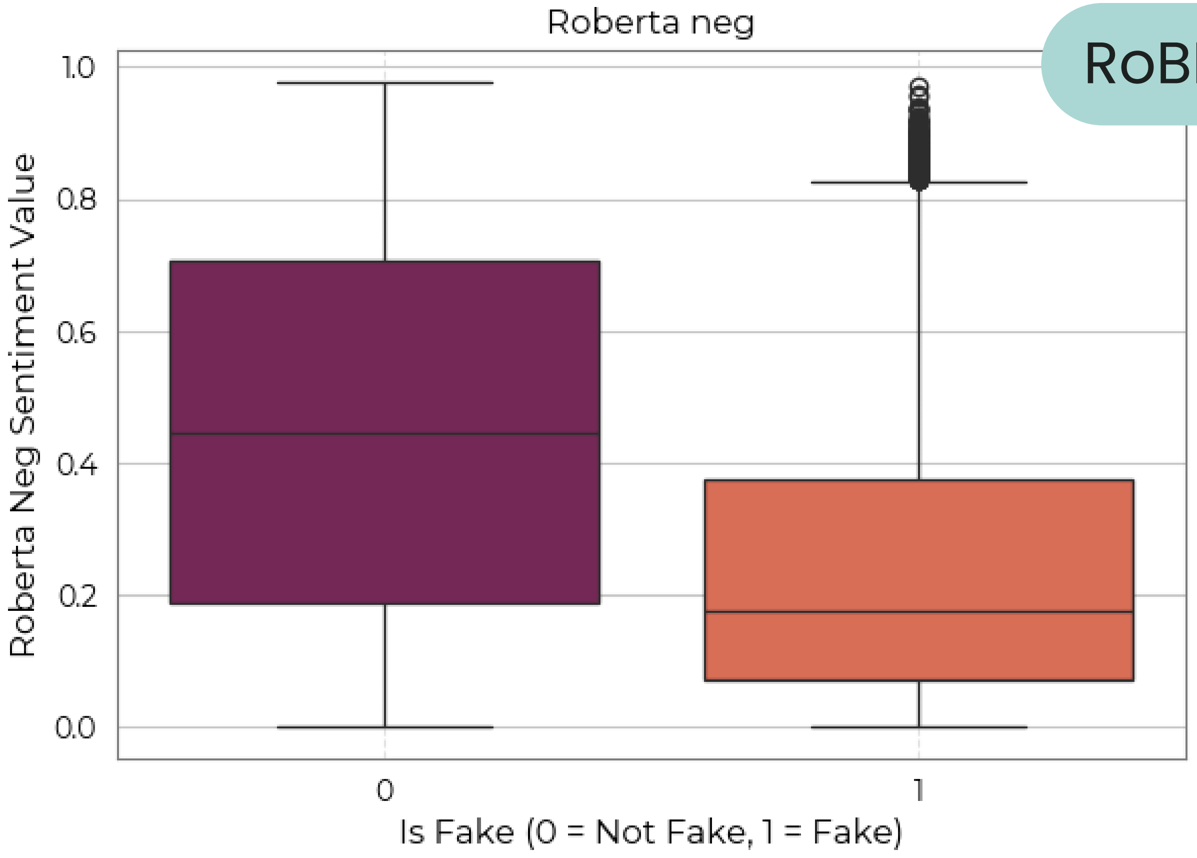


Comparison:

- While both models offer sentiment insights, RoBERTa is more effective at distinguishing between fake and non-fake headlines.
- This can be seen in the box plots, where the sentiment distribution for fake and not-fake headlines is more separated with RoBERTa than with VADER.

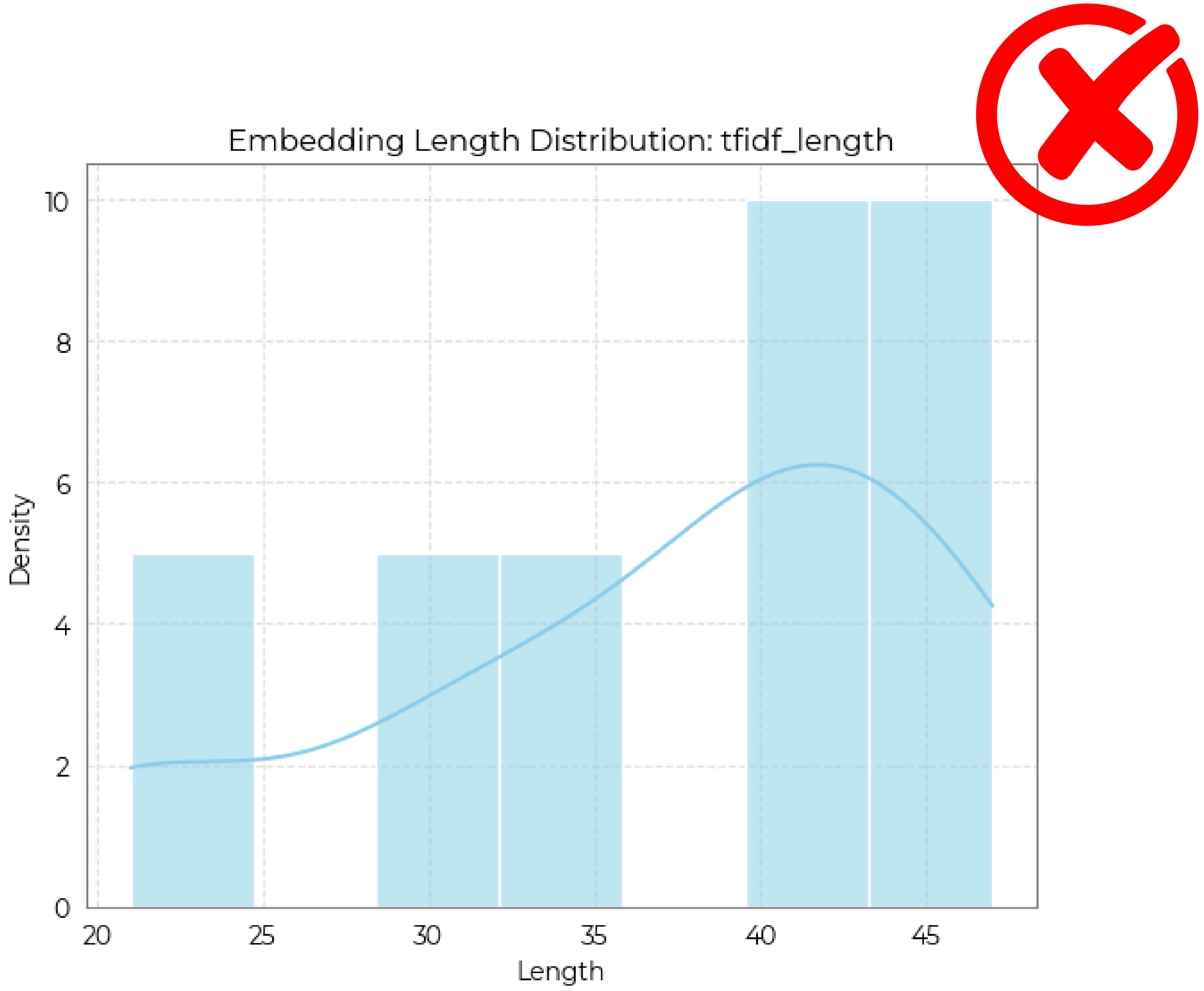
RoBERTa



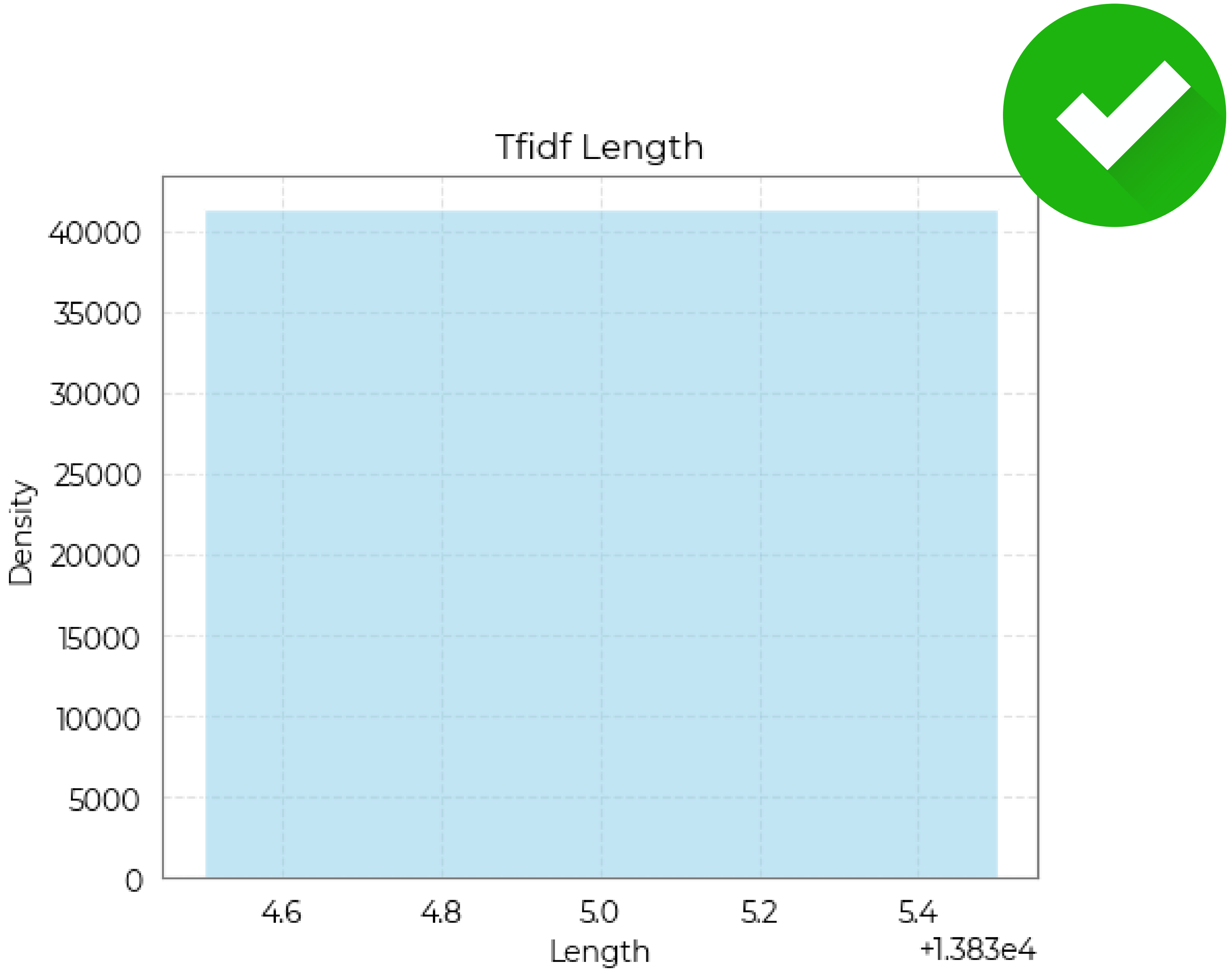


RobERTa has its own tokenization processes, stop word removal was not apply because the models rely on context. Lemmatization was also not applied.

Vectorization



- When training with more than one data frame



- When considering only the training data frame

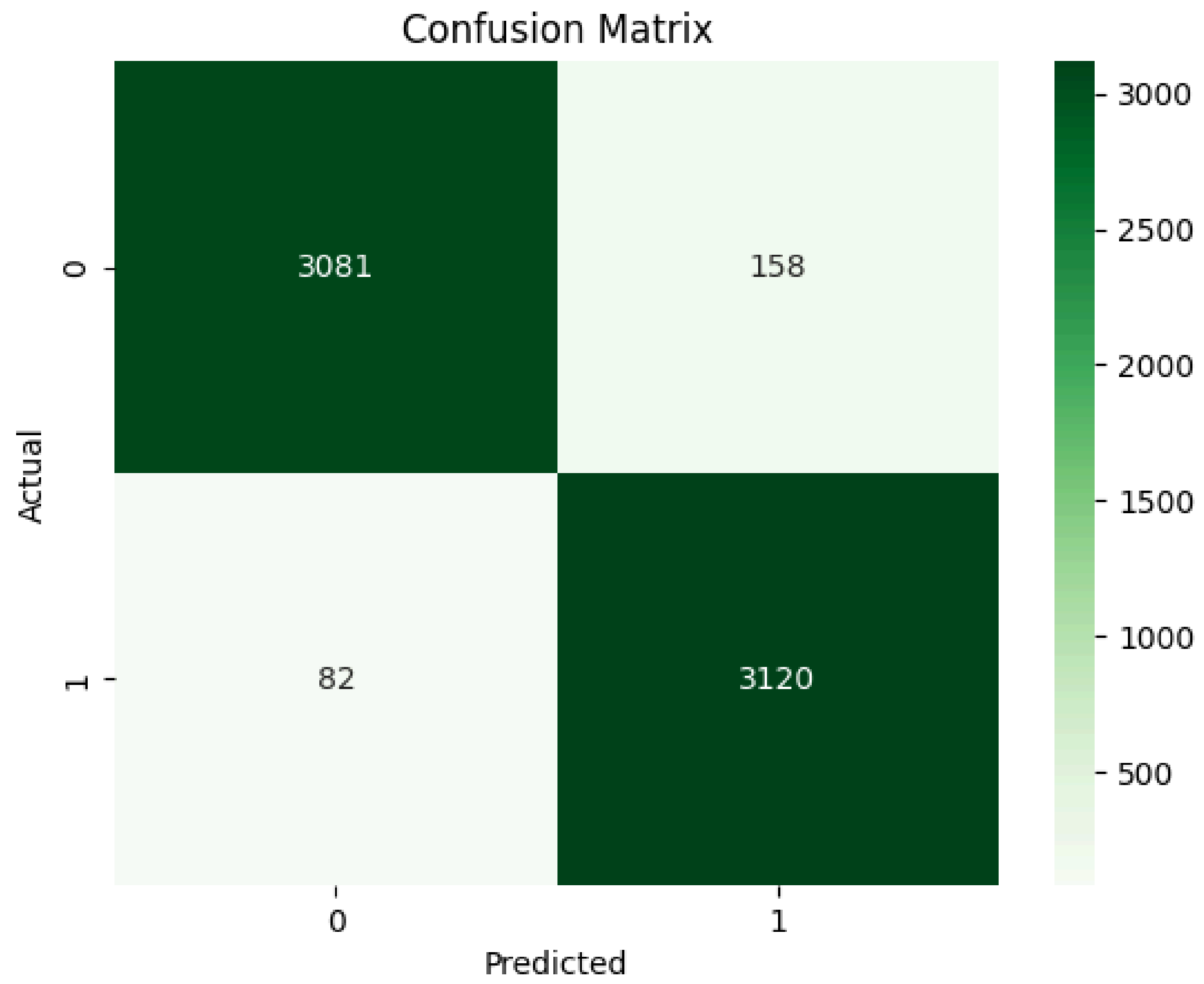
Model Training & Evaluation

Best model:

For all models: Flatten the vectors & StandardScaler()

Model	Accuracy	Precision	Recall	F1-score	Features
SVC	0.96	0.95	0.97	0.96	BERT embeddings alone

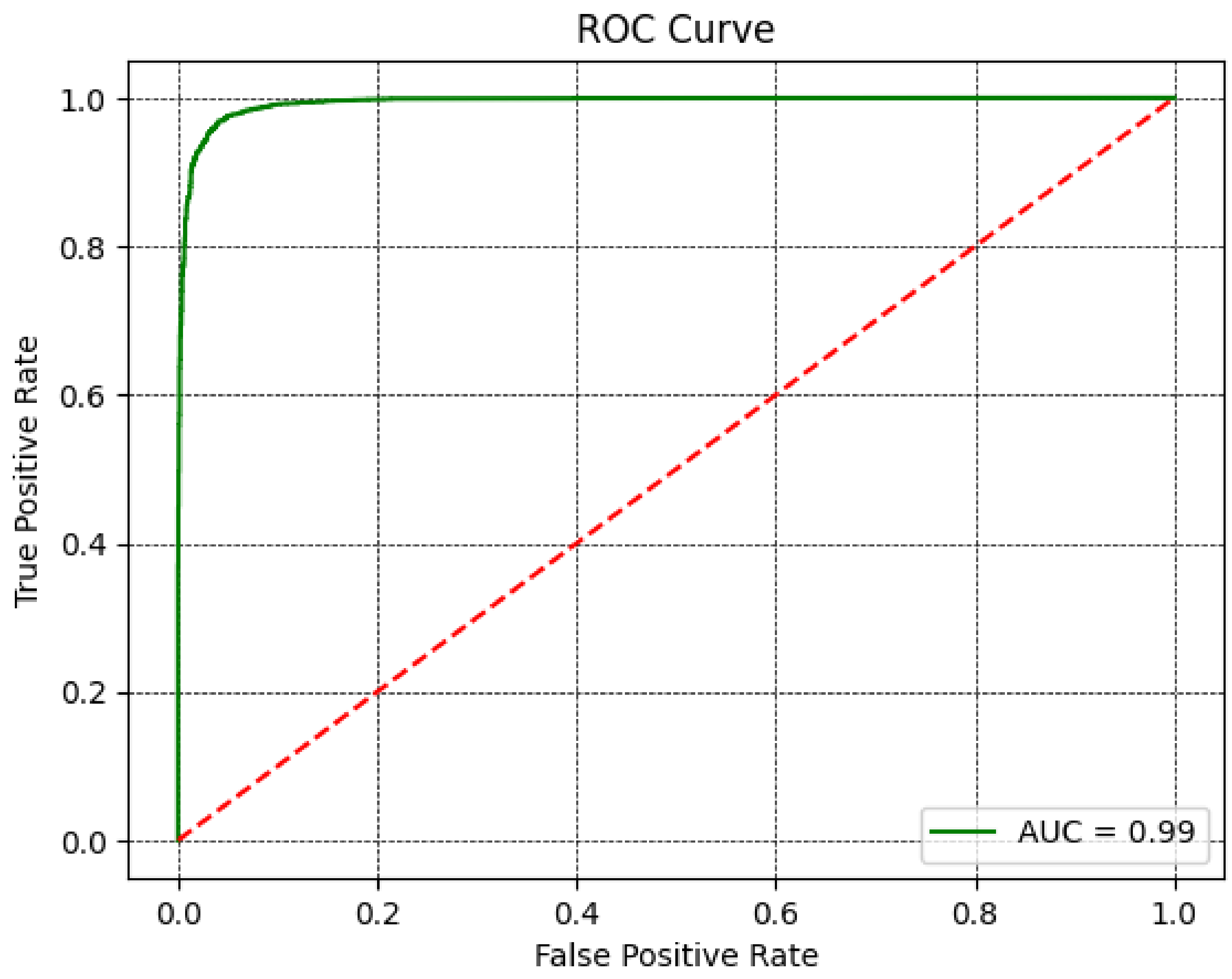
- **Accuracy for different vectors:**
  - Word Embeddings → **0.86** (with sentiments)
  - BoW → **0.91** (with or without sentiments)
  - BERT → **0.96** (with or without sentiments)
  - MiniLM → **0.95** (with or without sentiments)

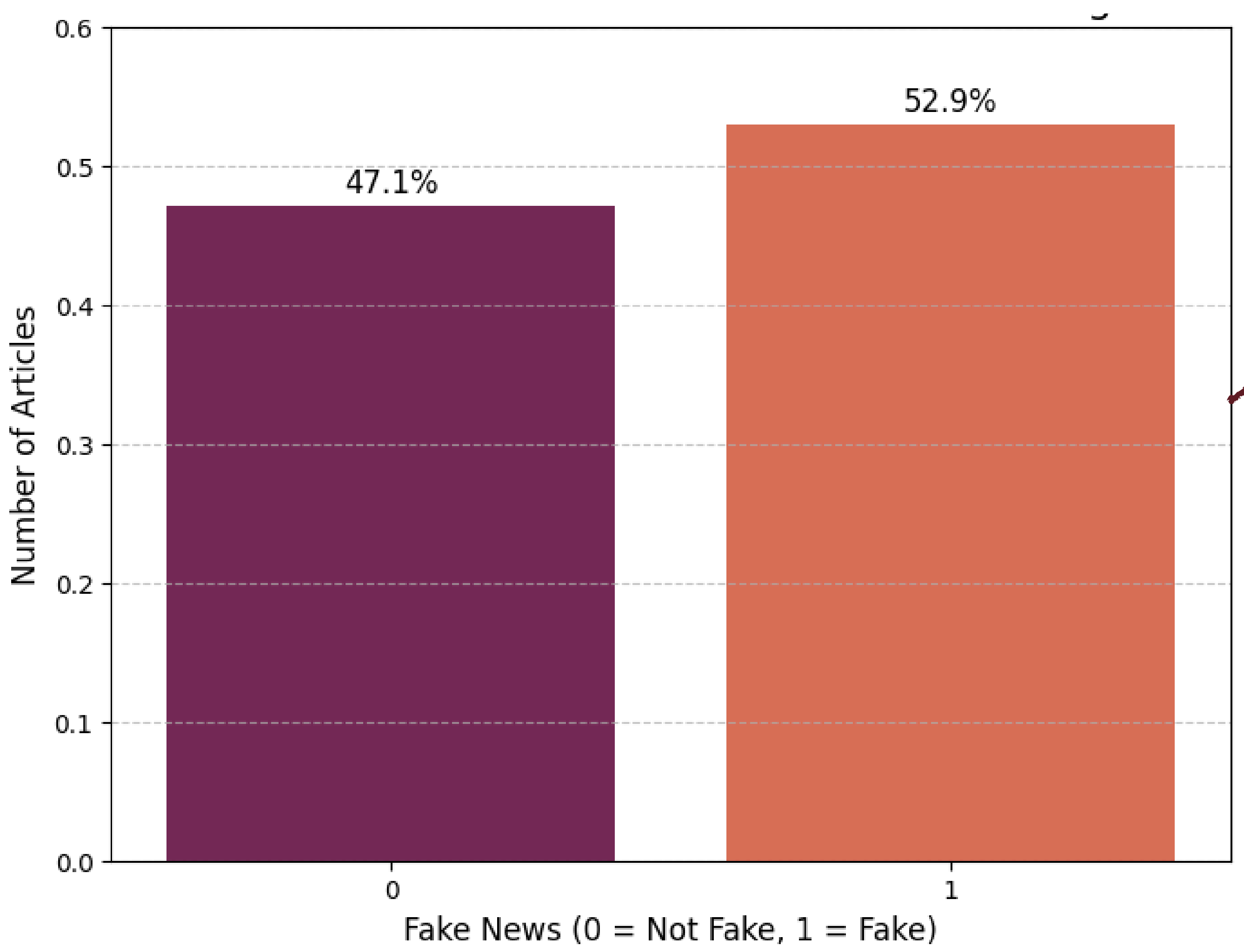


python + colab

Conclusions

- The model is highly effective for fake news detection, with strong performance across all key metrics (accuracy, precision, recall, F1-score).
- There is a minor trade-off between false positives and false negatives, but the overall performance suggests that the model is robust.
- Although the model performs well, there are still 158 false positives (non-fake news incorrectly labeled as fake). While this number is low, further improve the model might be possible by fine-tuning hyperparameters.





Relative distribution of Fake news articles in the testing dataset

Thank you very much!