Fake News Classification

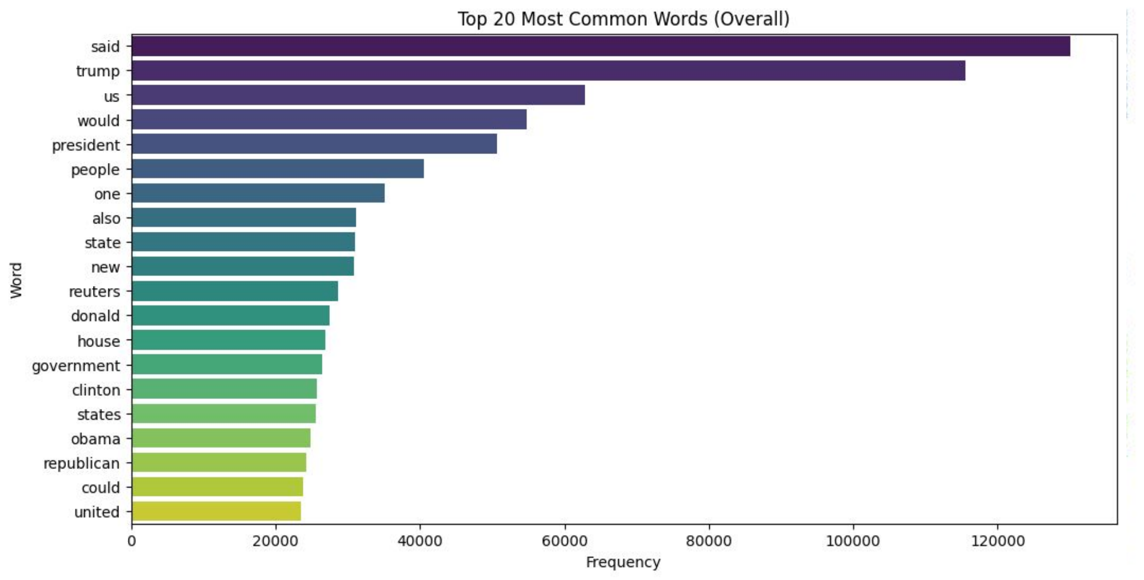
Kefan, Jiahao, Gavin, Tianyu, Kabir

Dataset Background

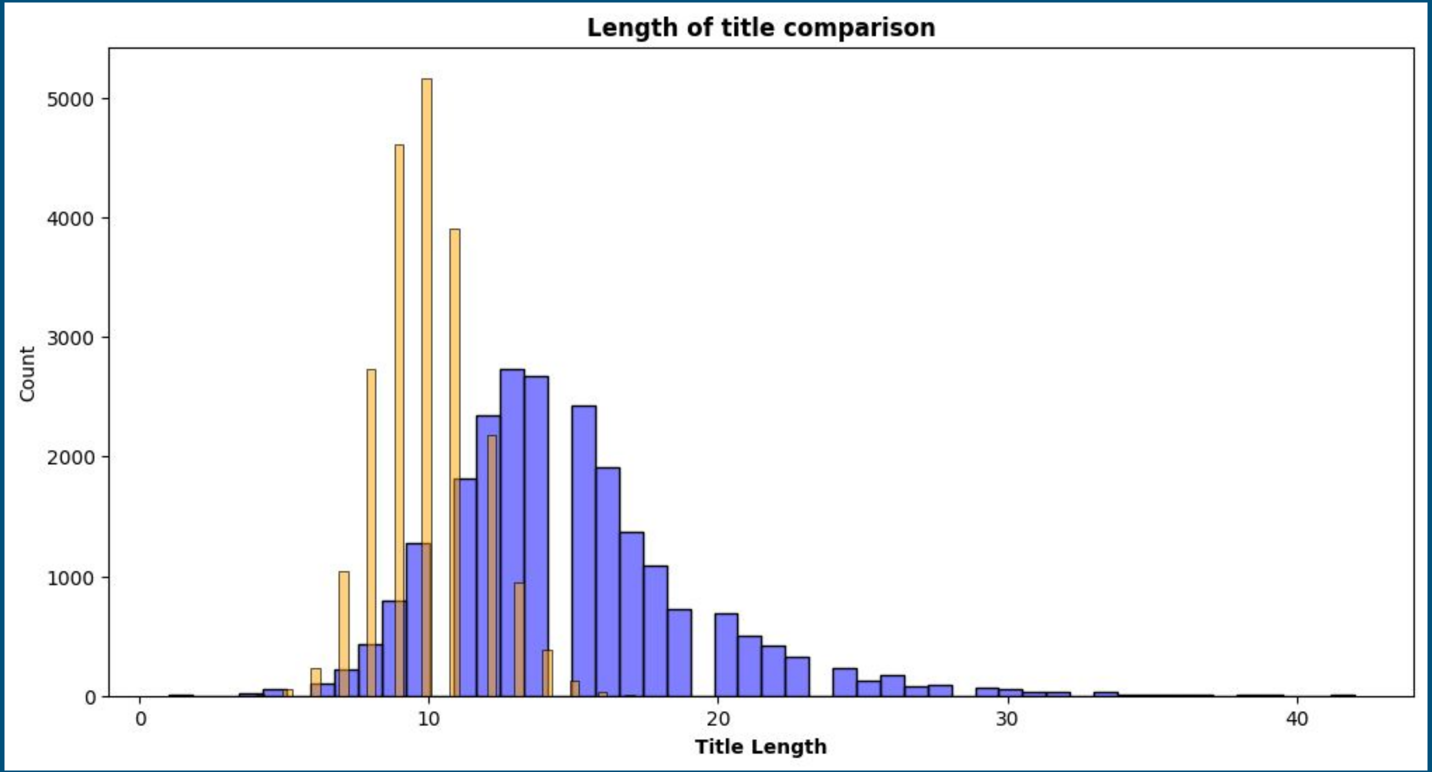
- https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/data
- Data from 2015–2017 news sources.
- The data is scraped from multiple sources:
    - Real news : From Reuters
    - Fake news : From 21st Century Wire, 100PercentFedUp, Twitter, …
- **NOT identifying truth,** Using features like writing style to identify whether from a **reputable** source

Exploratory Data Analysis

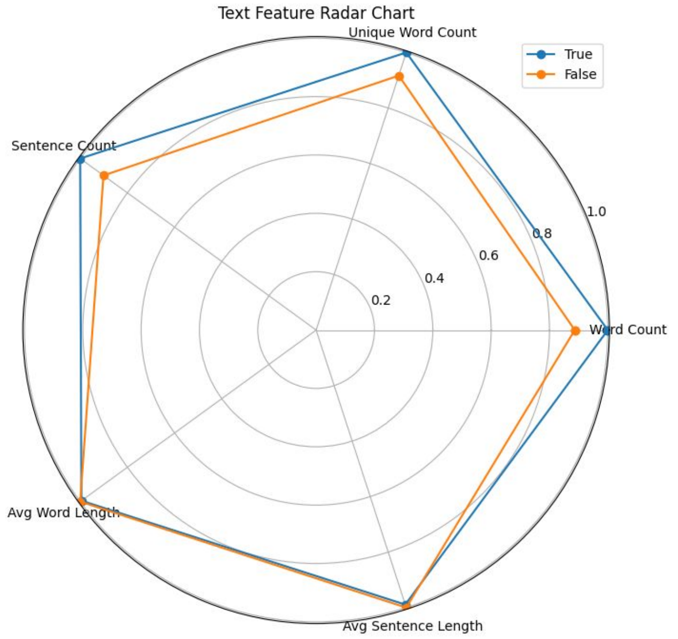
Primarily Composed of Political Discourse

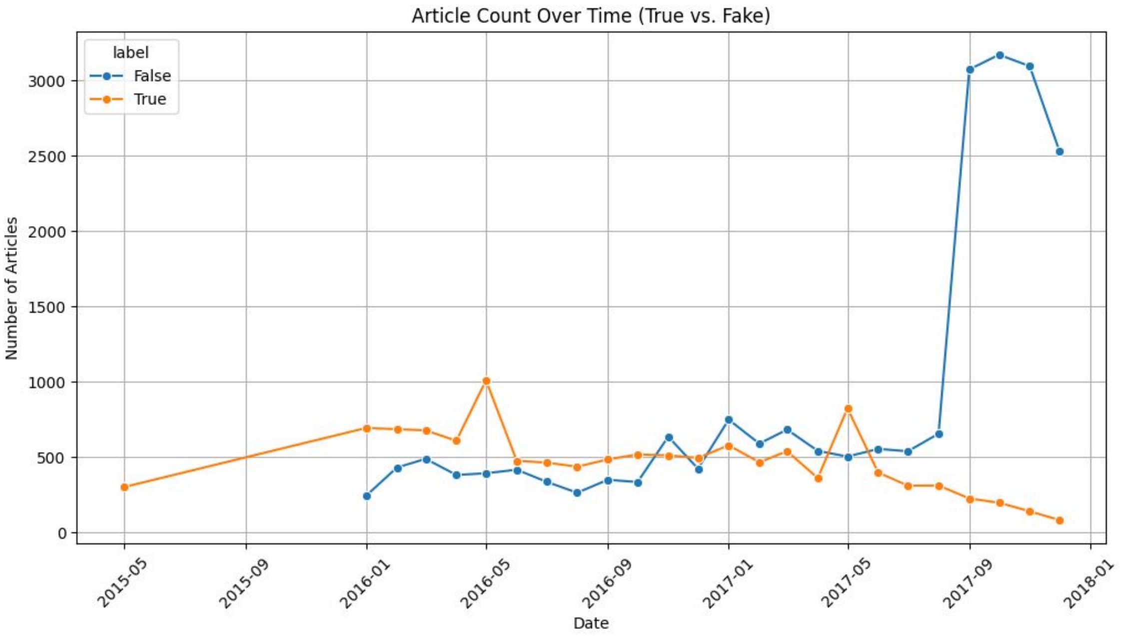


Data Analysis: title/text length comparison



Data Analysis: other features' comparison





Key Insights

Difference between True/False data -> Classification possible.

1. Different statistics and distributions of text

2. Different word frequency -> Different writing styles

3. Other features: Sentiment, …

Data Wrangling / Preprocessing

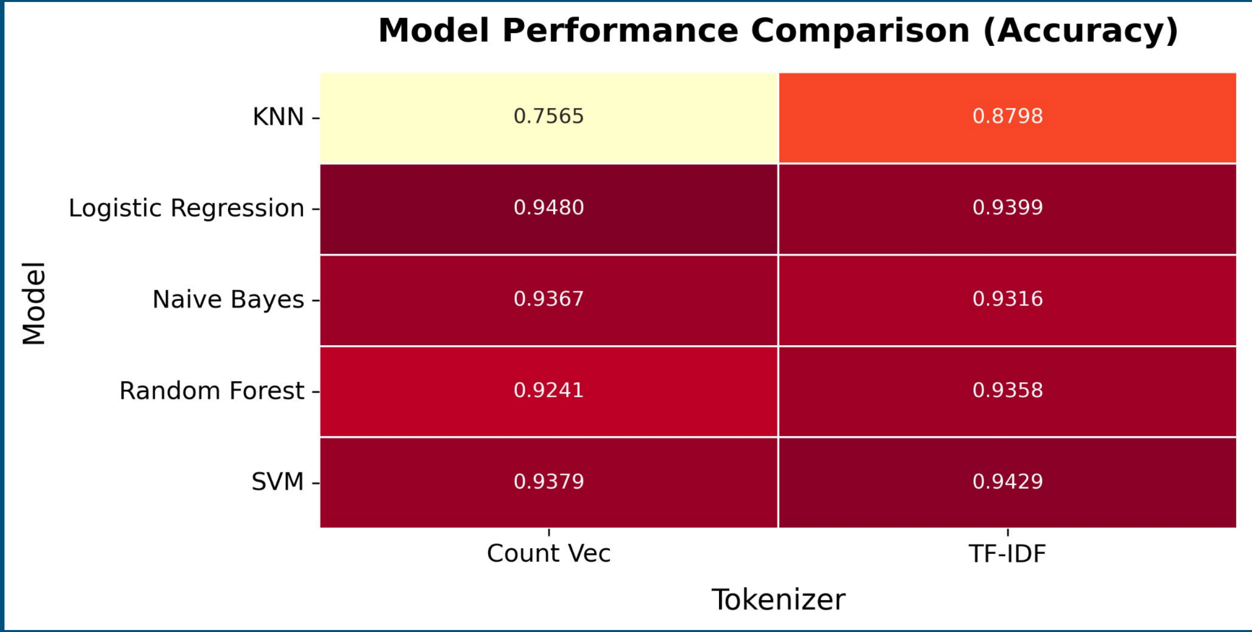
- Preprocessing Steps: Removed punctuation, spaces, and website links.
- Avoiding Overfitting: News source mentions (e.g., Reuters) led to 99% training accuracy but were removed, dropping accuracy to 50%.
- Key Insight: Titles provided more useful information for classification than full text.

Analysis Outline

- Used Traditional Machine Learning methods
- Used LSTM Recurrent Neural Network (Long Short Term Memory)
- Used BERT (transformer-based neural language model)
- Used LLM agents

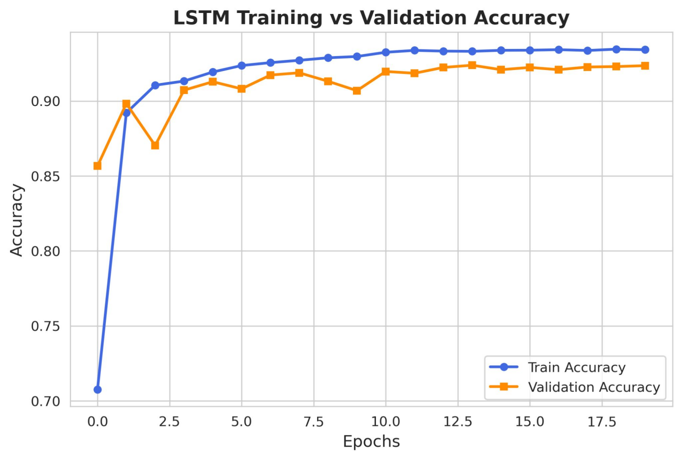
Traditional Machine Learning methods

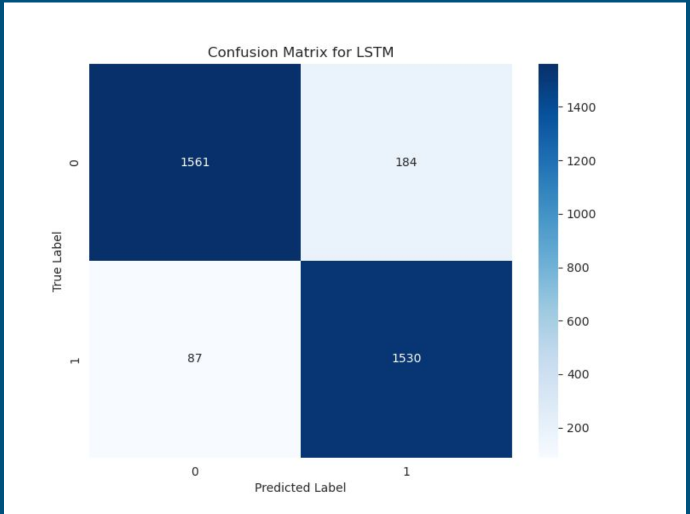
- Training on titles
- Acc lower than 50% with texts



Long Short Term Memory

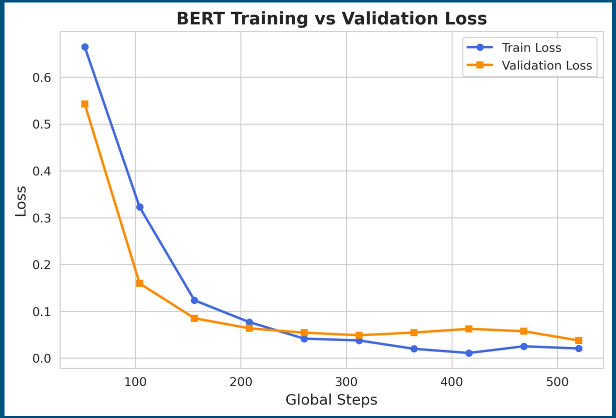
- LSTM 93% accuracy, in line with previous results
- LSTM's ability to predict well on just the titles as features again reinforces that the titles held all of the valuable information
- Expected higher results, but nearly identical to traditional ML

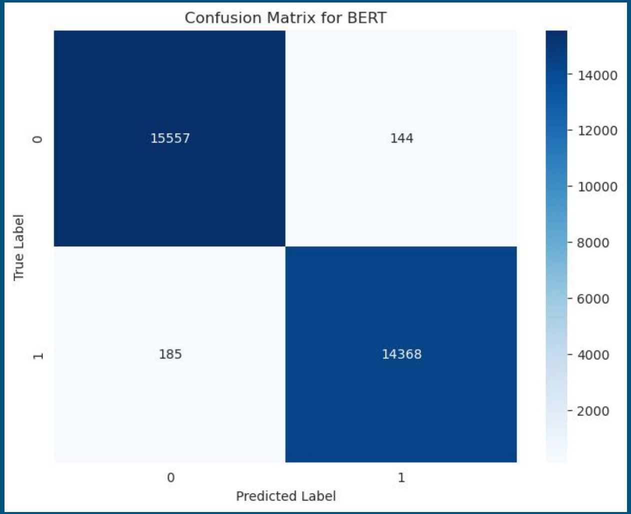




BERT Fine Tuning

- Training on just 10% of the data it achieved 99% accuracy classifying the remaining 90% test data
    - Feature: title + text
- BERT is REALLY good at identifying writing styles
- We couldn’t achieve a result higher than 50% with text alone - another evidence of titles having lots of info





Utilizing LLMs as News-Checker

Why LLMs?

- We want to see how much of the data labeled as “fake” is actually true
- Just because it was curated from these sources doesn’t mean it’s false. LLMs should know what did / didn’t happen between 2015-2017

Experiment Set-up

LLM APIs

- Deepseek-r1 provided by Deepseek
- Llama-3.3 provided by Groq

Prompt template

- Simple zero-shot prompt
- Include the title, text and date of a News

```
template="""
        You are a fact-checking assistant.

        Given the following news details:
        Title: {title}
        Text: {text}
        Date: {date}

        Determine if the news article appears more likely to
        be true, fake, or "NA" if policy limits your answer.

        Ensure that the output is a valid ptyhon dictionary.

        {format_instructions}
        """,
```

Experiment Set-up

Test data

- Randomly select 1000 samples from each .csv file

Output format

- A python dictionary which include:
- A “label”, either “true”, “fake”, “NA” or “Err”
- A”reason” than explain the label

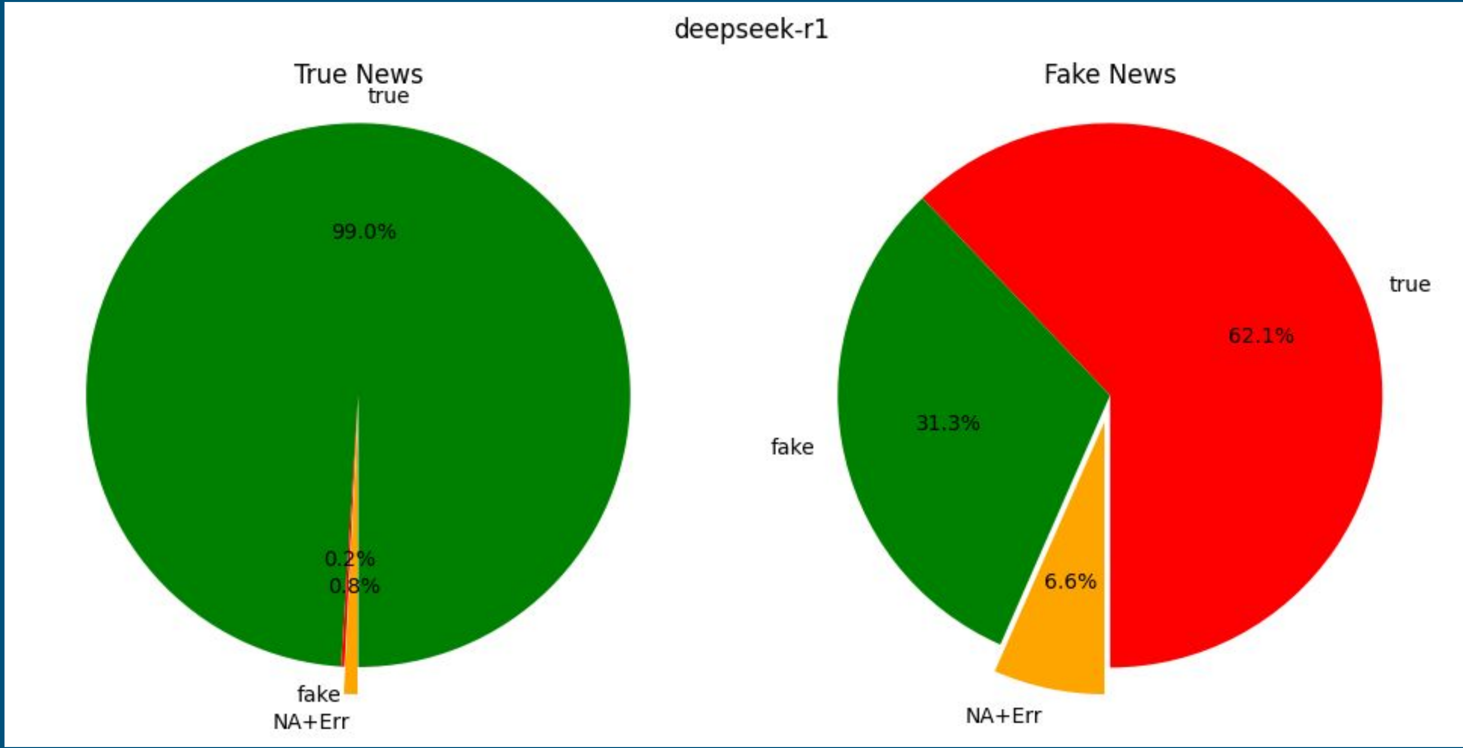
Input

```python
input = {
    'title': 'Scientists Invent Faster-than-Light Engine',
    'text': 'NASA declared that their scientists designed and tested the first '
    'faster-than-light engine in the world, which reached speeds up to three '
    'times the speed of light.',
    'date': 'March 10, 2024',
    'model': 'llama'
}
pprint.pprint(analyze_news(**input))
```

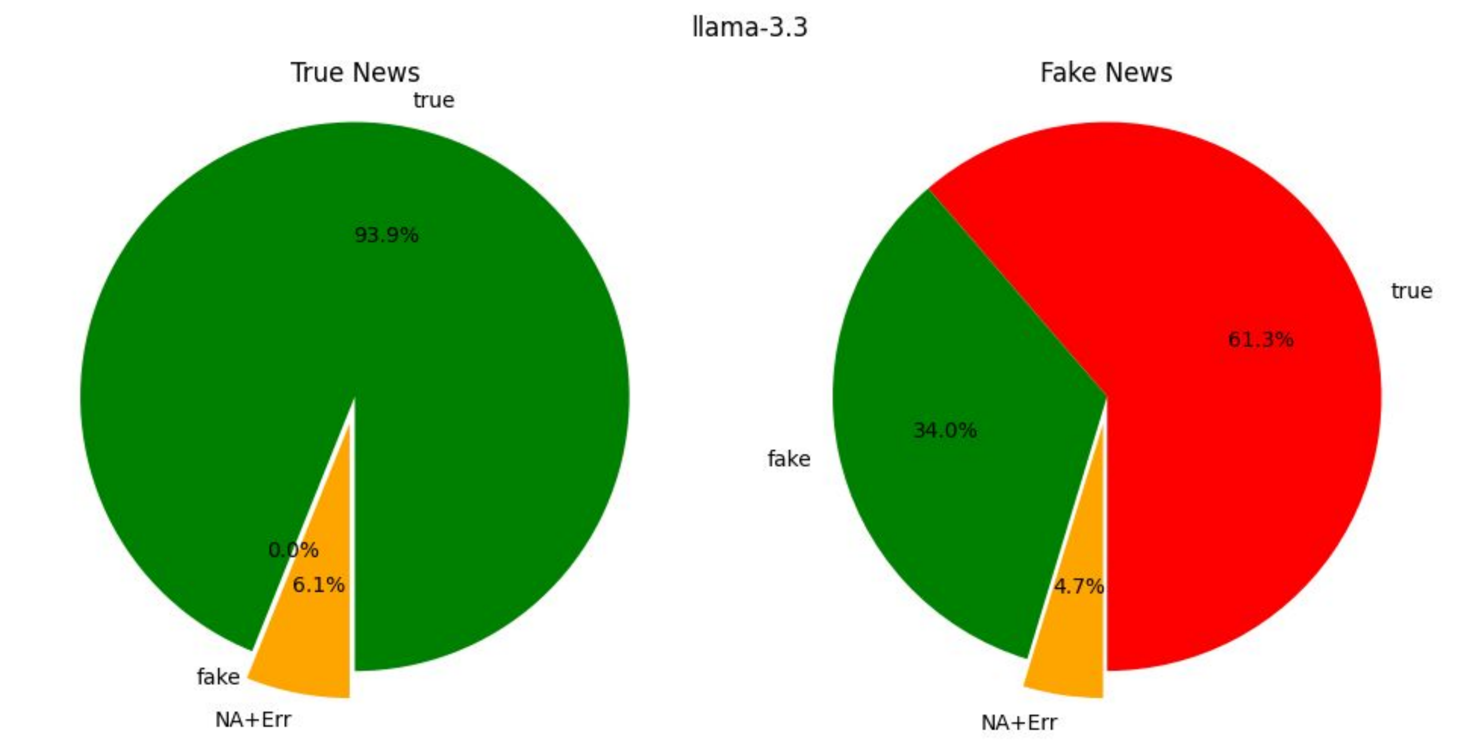
Output

```
{'label': 'fake',
 'reason': 'The claim of a faster-than-light engine contradicts the '
           'fundamental principles of physics, which state that the speed of '
           'light is the universal speed limit and cannot be exceeded by any '
           'object with mass.'}
```

Experiment Result



Experiment Result



Conclusions

- To identify if news comes from a reputable source, look at the title
    - The text body is much harder to make sense of (we couldn't)
- Traditional machine learning models can achieve high accuracy on identifying this from the title, same as LSTM
- BERT is *extremely* good at identifying writing source
- Not all disreputable news is false, but a large amount of it is