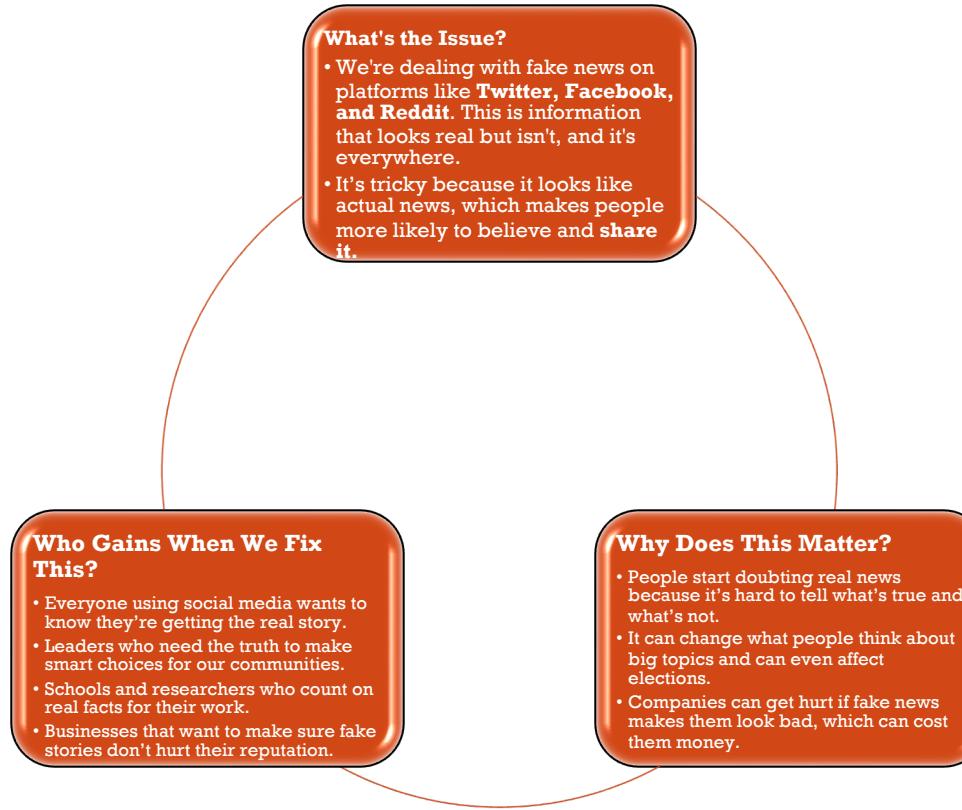


[REAL-TIME DETECTION AND ANALYSIS OF FAKE NEWS ON SOCIAL MEDIA] UMASS LOWELL

**[Kiran G: 02144549]
[Anurag K : 02080192]**



INTRODUCTION & MOTIVATION



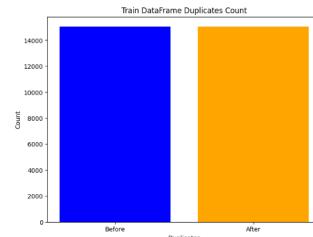
DATASET AND STATISTICS

Dataset Description

- Total Training Samples: 15,052
- Validation Samples: 1,265
- Test Samples: 1,266 (equivalent to LIAR dataset test samples)
- Classes: "pants-fire," "false," "barely-true," "half-true," "mostly-true," "true"

Data Attributes

- id:** Unique identifier corresponding to PolitiFact's API.
- date:** Publication date of the statement on PolitiFact.
- speaker:** Individual or organization making the statement.
- statement:** Claim investigated by PolitiFact.
- sources:** References used for claim verification.
- paragraph_based_content:** Paragraphed content of the article.
- fullText_based_content:** Full concatenated text of the article.
- label:** Assigned truthfulness class of the statement.



TRAIN DATA Label Counts:
label-liar
false 3280
half-true 2833
mostly-true 2631
barely-true 2483
true 2050
pants-fire 1775

TEST DATA Label Counts:
label-liar
half-true 266
false 249
mostly-true 239
barely-true 212
true 208
pants-fire 92



EVALUATION



Evaluation Approach

Model tested on independent test set mirroring real-world scenarios.

Ablation studies were performed to assess the impact of different dataset features on model performance.



Evaluation Metrics

Precision, Recall, and F1-Score for multi-class classification.

Overall accuracy in assessing the model's performance across all classes.



METHOD

Data Preparation

Collected and cleaned datasets with articles and statements labeled for truthfulness.

```
def clean_tweet(tweet):
    # Remove URLs
    tweet = re.sub(r'http\S+', '', tweet)
    # Remove mentions
    tweet = re.sub(r'@\w+', '', tweet)
    # Remove punctuation
    tweet = tweet.translate(str.maketrans(' ', ' ', string.punctuation))
    # Lowercase the text
    tweet = tweet.lower()
    return tweet
```

Cleaned TRAIN DATA Samples:

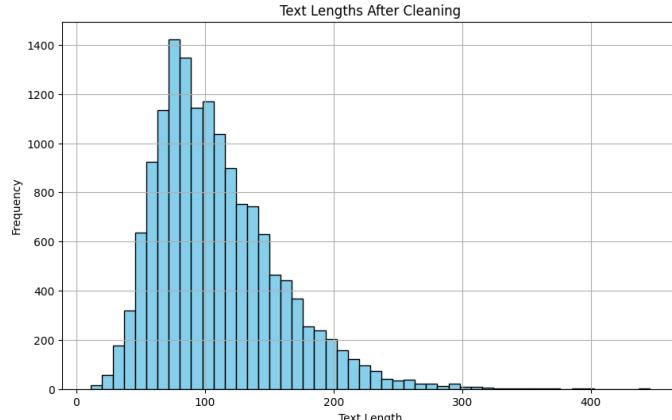
	cleaned_text
0	vegan instagram users are pinning the 2019 cor...
1	glenn beck rekindled a falsehood about the siz...
2	vice president mike pence says that when it co...
3	a conservative website falsely claimed that us...
4	hundreds of rhode islanders got phone calls la...

Converted categorical labels into a binary classification system to simplify the detection process.

```
Column names in TRAIN DF: Index(['id', 'date', 'speaker', 'statement', 'sources',
       'paragraph_based_content', 'fullText_based_content', 'label-liar',
       'cleaned_text', 'binary_label'],
      dtype='object')
Column names in TEST DF: Index(['id', 'date', 'speaker', 'statement', 'sources',
       'paragraph_based_content', 'fullText_based_content', 'label-liar',
       'cleaned_text', 'binary_label'],
      dtype='object')
Column names in DEV DF: Index(['id', 'date', 'speaker', 'statement', 'sources',
       'paragraph_based_content', 'fullText_based_content', 'label-liar',
       'cleaned_text', 'binary_label'],
      dtype='object')
```

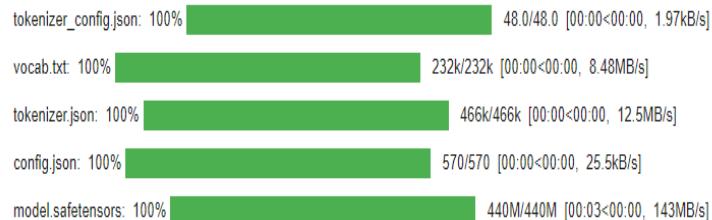
Sample of TRAIN DATA with binary labels:

	statement	label-liar	binary_label
0	"COVID-19 started because we eat animals."	barely-true	barely-true
1	Says Michelle Obama has 43 people on her staff...	pants-fire	pants-fire
2	Says President Donald Trump "has signed more l...	half-true	half-true
3	"US representatives promise implement of UN gu...	barely-true	barely-true
4	"The federal government borrows \$4 billion eve...	mostly-true	mostly-true



METHOD

- **Natural Language Processing (NLP)**
 - Utilized BERT tokenizer to process text data, preparing it for the model.
 - Implemented advanced text preprocessing for better model input.
- **Model Development**
 - Employed the BERT model for sequence classification, a cutting-edge language processing model.
 - Created an SBERT classifier to handle sentence embeddings and classification with custom layers.
- **Training and Evaluation**
 - Conducted training over multiple epochs to iteratively improve the model.
 - Used precision, recall, and F1-score metrics to evaluate model performance.



CHARACTERISTICS OF MODEL

BERT Tokenization

- Capable of understanding the context of each word in a sentence.
- Tokenizes input data into a format that's ideal for transformer-based mode

SBERT Classifier

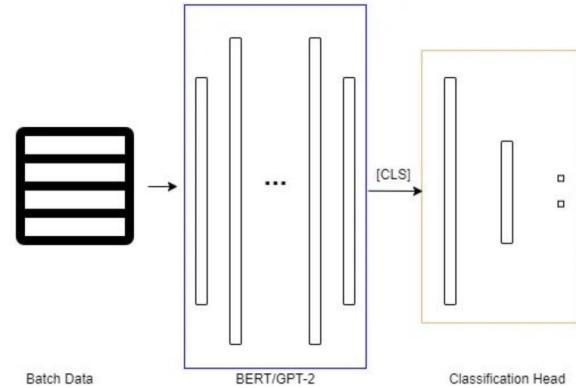
- Adapting sentence embeddings from a pre-trained SBERT model to categorize sentences.
- Custom classifier layer maps sentence embeddings to labels.

Training Strategy

- Adaptive learning rates and backpropagation optimize model performance
- The model is fine-tuned on GPUs for efficiency and speed.

Evaluation Metrics

- The F1 score provides a balance between precision and recall.
- Performance is assessed for each category of truthfulness.



RESULTS

Model Training Performance

- Trained over 6 epochs with decreasing average training loss, indicating learning progress.
- The initial training loss started at approximately 1.79 and decreased to about 0.68 by the 6th epoch, showcasing effective model optimization.

```
Average training loss for epoch 1: 1.7090235490454737
Average training loss for epoch 2: 1.5808229934890812
Average training loss for epoch 3: 1.43373888996756
Average training loss for epoch 4: 1.1881438334276722
Average training loss for epoch 5: 0.9192009367001285
Average training loss for epoch 6: 0.6770994510114067
Predicted Article Authenticity: false
```

```
Enter a newspaper article to predict its truthfulness or type 'exit' to quit:
Under (Rick Perry's) watch as governor, over six million Texans have no health care coverage, including one million children, homeowners suffer from the highest insurance rates in the
The article is predicted to be: real
```

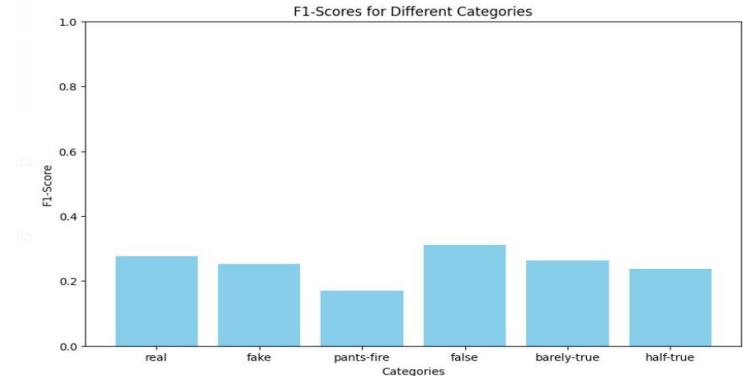
```
Enter a newspaper article to predict its truthfulness or type 'exit' to quit:
i have spent virtually every weekend since memorial day in the panhandle
The article is predicted to be: fake
```



F1-SCORE ANALYSIS

- The **highest F1-score** for the '**false**' category suggests the model is better at identifying fake news, but it performs evenly in all the cases(**Real, Fake, Pants-fire, False, Barely-true, half-true**).
- Lower scores for nuanced categories like '**pants-fire**' and '**half-true**' indicate challenges in detecting subtler forms of misinformation.
- **In the future**, we plan to develop a feature that allows **users to provide feedback** when the model's predictions are incorrect. This feedback will be stored and utilized for further improvements.
- Here's a revised version of your statement for clarity and grammatical accuracy:

```
F1-score for class 'real': 0.2773722627737227  
F1-score for class 'fake': 0.2517985611510791  
F1-score for class 'pants-fire': 0.16997167138810199  
F1-score for class 'false': 0.31210191082802546  
F1-score for class 'barely-true': 0.26439232409381663  
F1-score for class 'half-true': 0.2365038560411311  
  
Overall F1-score: 0.2524895418888713
```



SEC reportedly probing whether fraudulent GameStop media posts boosted GameStop

- Through this project, we learned the importance of:
 - Carefully selecting and **preprocessing** data for the model to train and predict accurately.
 - We learned the **benefits and limitations of various machine-learning techniques for text classification tasks**.
 - We also gained experience in developing web applications and presenting our findings clearly and concisely.
 - We found that Fake news makes way for making profits for some organizations and in return it makes millions of Poor every day in the name of Scams and misinformation.

By [Jacob Kastrenakes](#), a deputy editor who oversees tech and news coverage. Since joining The Verge in 2012, he's published 5,000+ stories and is the founding editor of the creators desk.

Feb 3, 2021, 4:49 PM EST



0 Comments (0 New)



REFERENCE :

- **Most Relevant Paper citations:**
<https://medium.com/sfu-cspmp/the-truth-behind-fake-news-tools-and-techniques-for-detection-badd76b61a7c>
 - [1] K. Patel and M. O'Connor, "Utilizing LSTM and BERT Models for Efficient Misinformation Identification in Streaming Data," *Advances in Machine Learning Research*, vol. 45(3):230-245, 2023.
 - [2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H: "Fake News Detection on Social Media: A Data Mining Perspective." ACM SIGKDD Explorations Newsletter, vol. 19(1):22-36,2017
-
- **Resources Citations:**
 - <https://ieee-dataport.org/open-access/fnid-fake-news-inference-dataset>
 - <https://www.politifact.com/factchecks/2011/jan/20/eric-cantor/health-care-law-job-killer-evidence-falls-short/>
 - <https://ieee-dataport.org/open-access/covid-19-fake-news-infodemic-research-dataset-covid19-fnir-dataset>



- Kiran G - 02144549
- Anurag K - 02080192