# Fake News Detection Using Machine Learning Models


*Elise Rust*

*Georgetown University*

*Computational Linguistics Spring 2022*


GEORGETOWN UNIVERSITY

In the era of disinformation, fraudulent text propagated by major political figures, news outlets, and social media platforms is challenging to identify. This project studies how different ML models and NLP pre-processing perform in fake news classification.

## Background

- [1] Comparative survey of ML methods in fake news detection: **SVM, NB, Log Reg, DT, NN, CNN, and RST**
- [2] Inclusion of RF model for 73% accuracy
- [3] Inclusion of **NER** to calculate semantic features = 5-10% accuracy improvement

| Model | Metric Name | Metric Value |
|---|---|---|
| Bi-LSTMs | Test Accuracy | 0.223 |
| CNNs | Test Accuracy | 0.27 |
| CNNS: Text + Speaker | Test Accuracy | 0.248 |
| Text + All | Test Accuracy | 0.274 |

## Data

- **LIAR dataset** (Wang et. al 2017)
- Multiclass labels: ["True, Half-True, Mostly-True, Barely-True, False, Pants-Fire"]
- Speaker, Subject Matter, Political Affiliation
- 10 years, **12.8K manually labeled** statements from PolitiFact.com


**References**

[1] *Oshikawa, Ray, Jing Qian, and William Wang. "A survey on natural language processing for fake news detection."(2018).*

[2] *Khanam, Z., B. N. Alwasel, H. Sirafi, and M. Rashid. "Fake news detection using machine learning approaches." (2021)*

[3] *Brasoveanu, Adrian MP, and Râzvan Andonie. "Semantic fake news detection: a machine learning perspective."(2019)*

[4] Wang, William Yang. "" liar, liar pants on fire": A new benchmark dataset for fake news detection." (2017)


## Methodology

- **TFIDF-vectors of text data**
  - Text Pre-Processing
    - Stemming
    - Remove punctuation, stopwords
    - Replace numbers w/ '#'
- Combined with Speaker, Subject, Party data
- 3 ML models
  - Support Vector Machine (SVM)
    - Linear and Polynomial Kernels
  - Naïve Bayes (NB)
    - Multinomial NB
  - Decision Trees
- Calculate Success:
  - **Accuracy, Precision, Recall, F1 Score**
- Mutual Information Analysis

## Mutual Information Analysis

*What can one variable tell us about the validity of a statement? What keywords are key to classification?*

**Table 3 – Mutual Information**

| | Top Values | Max MIC |
|---|---|---|
| Text | would cut, compens, almost, item, Toomey, billion, import, proven, ben, say | 0.022 |
| Speaker | ami, newsmax, rooney, ahern, gavin, disease, burnt, olson, sarah, leeder | **0.027** |
| Subject | energy, prices, transportation, islam, lottery, drugs, justice, military, regulation, occupy | **0.027** |
| Political Party | body, business, liberal, moderate, democrat, activist, show, state, talk, farmer | 0.024 |

- 'Speaker' and 'Subject' tied for highest MIC score **(0.027)**
- 'Subject' important keywords = hot button topics
  - i.e. *energy, prices, justice, military, drugs*

Code: https://github.com/eliserust/Fraudulent-Text-Detection

## Multiclass Results

**Best Predictive Model = Support Vector Machines**

**Table 1 – Multiclass Accuracy of Different Models**

| | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| **No Pre-Processing** | | | | |
| SVM - Linear | 0.235 | 0.227 | **0.225** | **0.22[illegible]** |
| SVM - Polynomial | **0.253** | **0.405** | 0.216 | 0.18[illegible] |
| Multinomial NB | 0.243 | 0.206 | 0.209 | 0.19[illegible] |
| Decision Tree | 0.224 | 0.249 | 0.194 | 0.16[illegible] |
| **Text Pre-Processing** | | | | |
| SVM – Linear | 0.231 | 0.223 | **0.221** | **0.22[illegible]** |
| SVM – Polynomial | 0.242 | **0.428** | 0.204 | 0.17[illegible] |
| Multinomial NB | **0.246** | 0.266 | 0.215 | 0.20[illegible] |
| Decision Tree | 0.210 | 0.229 | 0.189 | 0.15[illegible] |

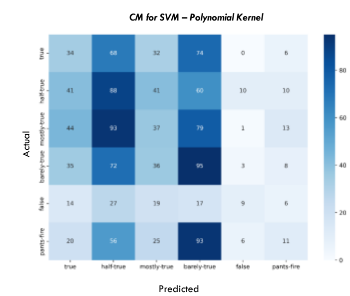


**Table 2 – Ablation Table for Preprocessed Text**
*(Accuracy of Multiclass Models)*

| | Linear SVM | Poly SVM | Multinomial NB | DT |
|---|---|---|---|---|
| Just Text | 0.231 | **0.242** | 0.246 | 0.210 |
| Just Subject | 0.214 | 0.211 | 0.204 | 0.207 |
| Just Speaker | 0.215 | 0.235 | 0.211 | 0.231 |
| Just Political Party | 0.222 | 0.222 | 0.217 | 0.223 |
| Text + Subject | 0.239 | 0.192 | 0.251 | 0.231 |
| Text + All | **0.244** | 0.225 | **0.256** | **0.239** |

- Inconsistent Performance
- Low predictive power
  - 21-25% accuracy
  - 22 – 42% precision
- Inclusion of Speaker, Party, Subject
  - ↑ of 0.008
- Commonly mislabeled:
  - True → Barely True
  - Half-True → Barely True
  - Mostly True → Barely True
  - Barely True → Half True
  - False → Half True
  - Pants Fire → Barely True

## Discussion

- Context is critical: Speaker, Party affiliation, Subject
- Performed similar to baseline (25.6% accurate)
  - Random Forest, CNN, NN, RST worth exploring
- Ongoing research needed to address **fuzzy class boundaries**
  - Distinctions between "Half-true", "Barely-True" "Mostly-True" are UNCLEAR and unscalable