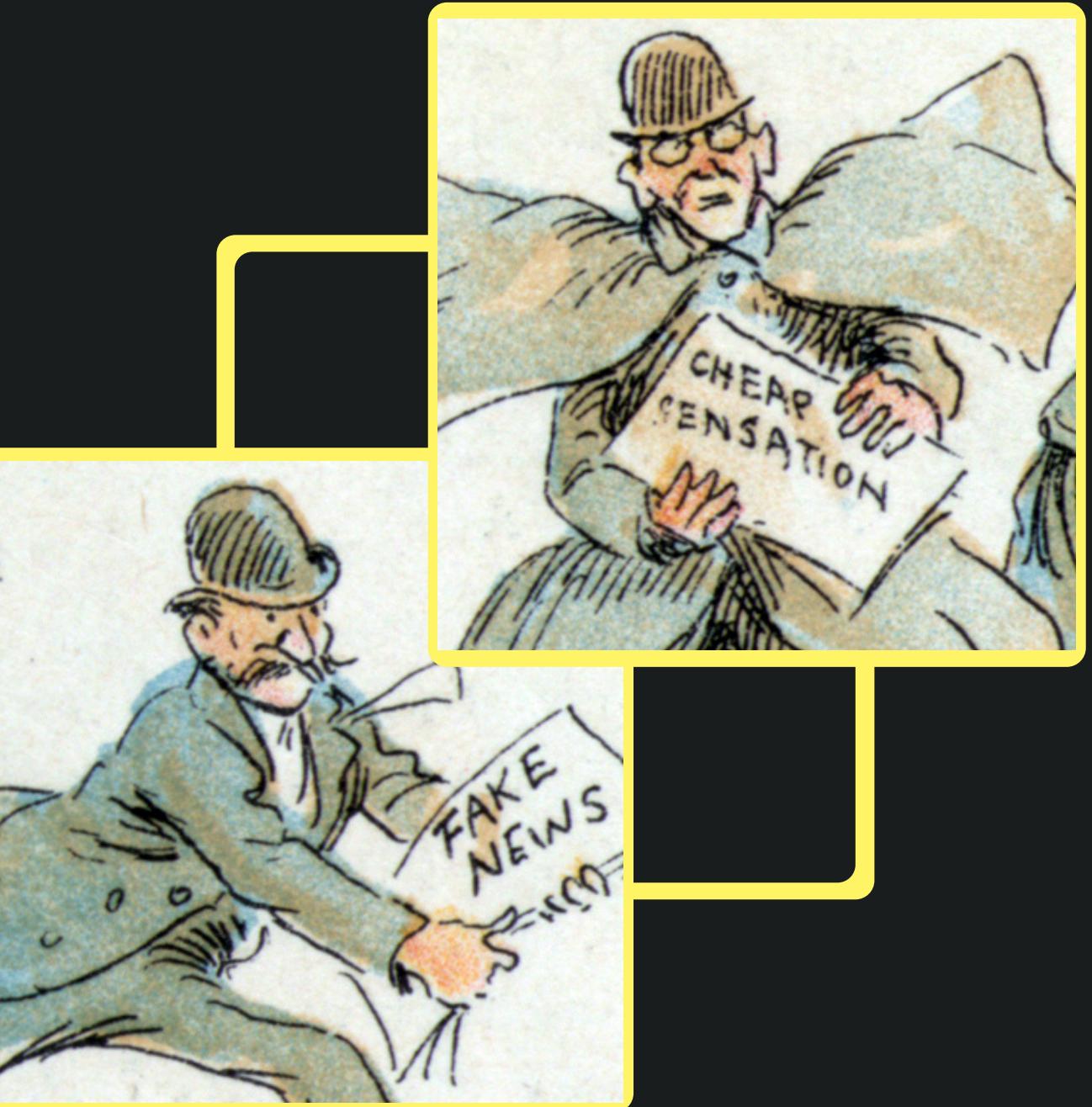


# Classifying Fake News

By Marcelo Scatena



# Overview

01 Business Problem

02 Data & Methods

03 Modeling

04 Conclusion



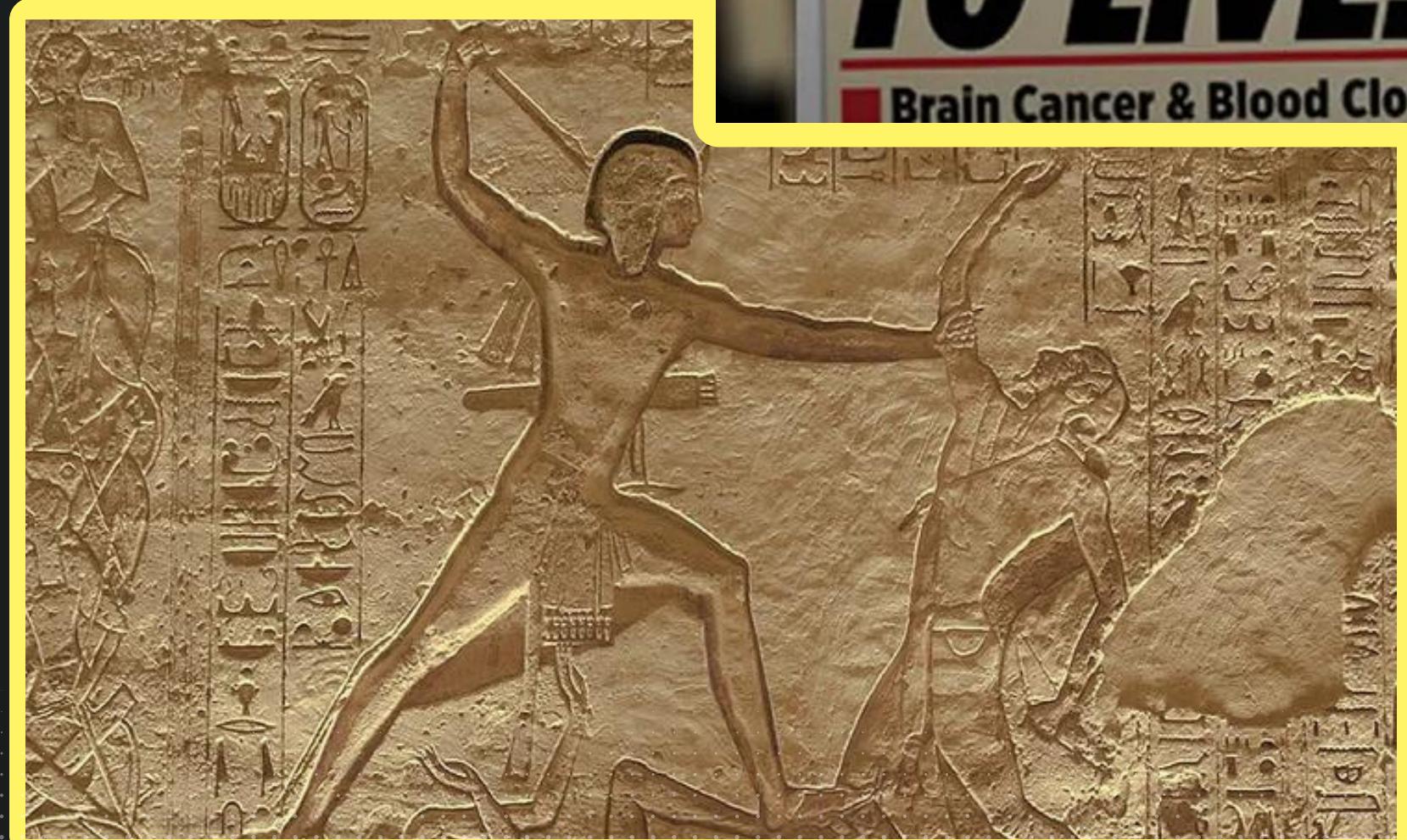
# Business Problem

Has been around for a long time

- 2016 US Presidential Election
- WWII Propaganda
- 13th bc Battle of Kadesh

Can have many different types

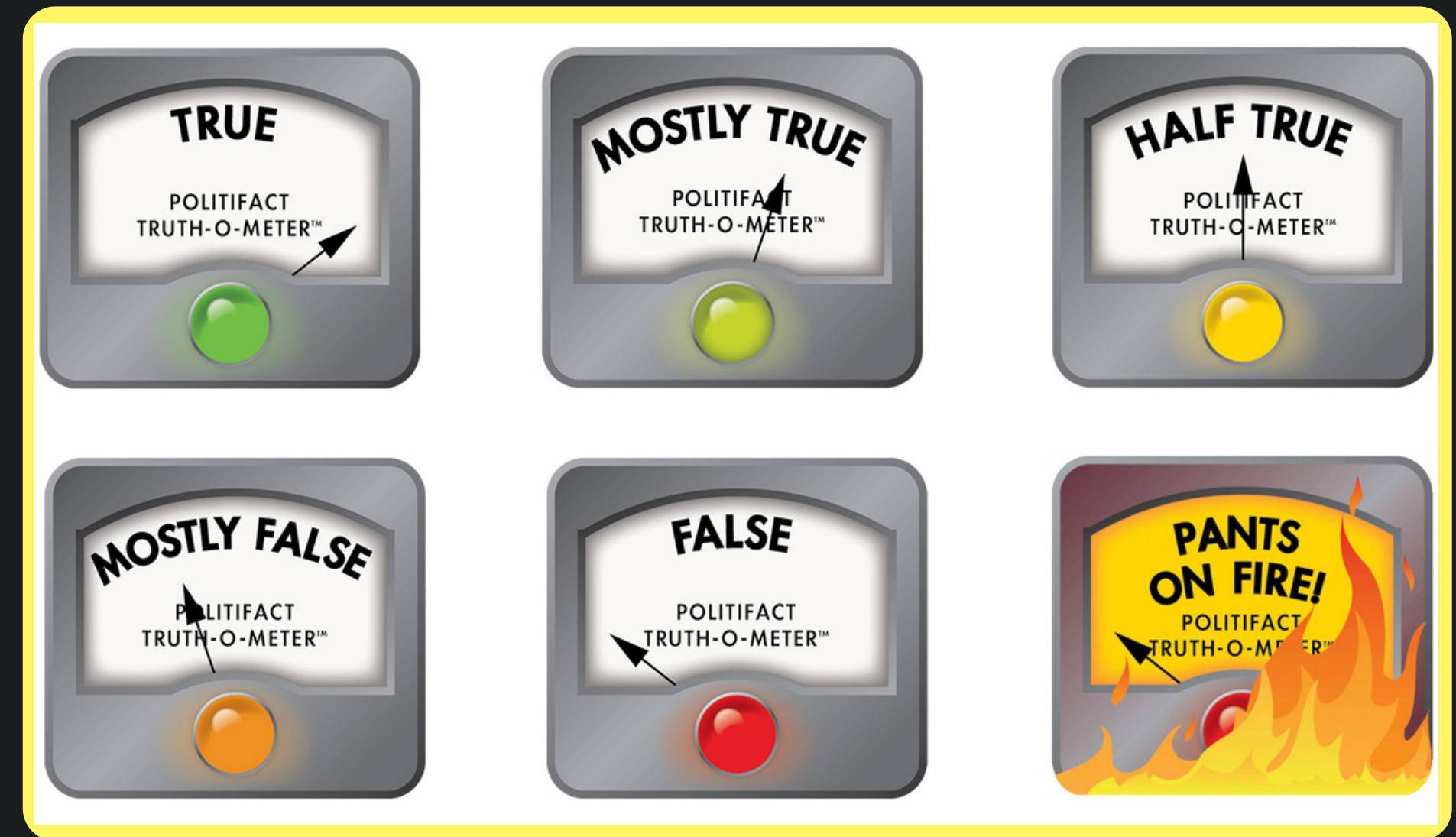
- Fabricated Content
- Manipulated Content
- Parody
- False Connection
- Manipulated Content
- Misleading Content
- Impostor Content



# Data & Methods

## LIAR dataset

- 12.8 thousand statements
- Collected with a 10 year range
- Humanly labelled in 6 categories
- 12 columns of metadata



# Politifact .com



Tweets stated on March 01, 2022 in :

**Video shows Ukrainian crisis actors come to life after pretending to be dead.**

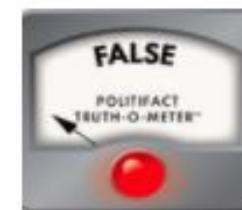
By Samantha Puttermann • March 4, 2022



Facebook posts stated on February 28, 2022 in :

**A Ukrainian TikToker shows how to drive an abandoned Russian armored vehicle.**

By Samantha Puttermann • March 4, 2022



Facebook posts stated on February 25, 2022 in :

**The war in Ukraine is “scripted and staged.”**

By Gabrielle Settles • March 4, 2022



# Preprocessing

Original Statement

**"Twitter, Google and Facebook are burying the FBI  
criminal investigation of Clinton. Very dishonest media!"**

Split words,  
remove punctuation &  
lowercase

**twitter, google, and, facebook, are, burying, the, fbi,  
criminal, investigation, of, clinton, very, dishonest, media**

Remove Stopwords

**twitter, google, facebook, burying, fbi, criminal,  
investigation, clinton, dishonest, media**



# Preprocessing

## Categorical Encoding

**twitter, google, facebook, burying, fbi, criminal,  
investigation, clinton, dishonest, media**

twitter: 0

google: 1

facebook: 2

burying: 3

...

media: 9



# Preprocessing

## Categorical Encoding

**twitter, google, facebook, burying, fbi, criminal,  
investigation, clinton, dishonest, media**

twitter: 0  
google: 1  
facebook: 2  
burying: 3  
...  
media: 9



[0, 1, 2, 3...9]



# Preprocessing

## Categorical Encoding

twitter, google, facebook, burying, fbi, criminal,  
investigation, clinton, dishonest, media

twitter: 0  
google: 1  
facebook: 2  
**burying: 3**  
...  
media: 9



[0, 1, 2, **3**...9]



[1, 0, 0, 0, 0, 0, 0, 0, 0]  
[0, 1, 0, 0, 0, 0, 0, 0, 0]  
**[0, 0, 1, 0, 0, 0, 0, 0, 0]**  
...  
[0, 0, 0, 0, 0, 0, 0, 0, 1]



# Preprocessing

## Word Embedding

- Word2Vec



# Preprocessing

## Word Embedding

- Word2Vec



**health**

```
array([-0.95737183,  0.30383423,  0.36267897, -0.3145774 , -0.40760833,
       -0.9460593 ,  0.49187672,  1.4210457 , -0.30092633, -0.20819204,
       -0.770027 , -0.7723512 , -0.33388203, -0.1773534 ,  0.20409673,
       -0.5343583 , -0.39575577, -0.93514854, -0.06234565, -1.7523842 ,
       0.22604117,  0.65160227,  0.5052533 , -0.26991683, -0.3610879 ,
       -0.53496146, -0.32183996, -0.92692065,  0.07382824,  0.26401505,
       0.7777989 ,  0.04584725,  0.21644872, -0.6334908 ,  0.10794742,
       1.0837792 ,  0.05388981,  0.19897468,  0.6462655 , -1.1305623 ,
       0.15471256, -0.7179126 , -0.15460503, -0.32296193,  0.1598994 ,
       -0.5075027 , -0.7410554 ,  0.5247071 ,  0.0338988 ,  0.06148871,
       -0.26969084, -0.40099415,  0.30752096,  0.11141815, -0.5975455 ,
       0.9567183 ,  0.71071416, -0.3935786 , -0.60557264,  0.17251973,
       0.42103177,  0.02287591,  0.02722805, -0.9834109 , -0.35429567,
       0.23459306,  0.19908333,  0.42050263, -0.80189645,  0.11037309,
       -0.26703724,  0.9878612 ,  0.63543224, -0.02554742,  0.96865296,
       0.15772453,  0.29907194, -0.0074525 , -1.3398806 ,  0.17438231,
       -0.1138156 ,  0.4435544 , -0.30574077,  0.52437586,  0.2716861 ,
       -0.13374457,  0.034655 ,  0.8558606 ,  0.20205818, -0.17715032,
       0.5058484 ,  0.09029104,  0.16534522, -0.28038573,  1.0729178 ,
       1.0143198 , -0.3190712 , -0.8396994 ,  0.5295455 ,  0.6986663 ],
      dtype=float32)
```



# Preprocessing

## Word Embedding

- Word2Vec



health

```
array([-0.95737183,  0.30383423,  0.36267897, -0.3145774 , -0.40760833,
       -0.9460593 ,  0.49187672,  1.4210457 , -0.30092633, -0.20819204,
       -0.770027 , -0.7723512 , -0.33388203, -0.1773534 ,  0.20409673,
       -0.5343583 , -0.39575577, -0.93514854, -0.06234565, -1.7523842 ,
       0.22604117,  0.65160227,  0.5052533 , -0.26991683, -0.3610879 ,
       -0.53496146, -0.32183996, -0.92692065,  0.07382824,  0.26401505,
       0.7777989 ,  0.04584725,  0.21644872, -0.6334908 ,  0.10794742,
       1.0837792 ,  0.05388981,  0.19897468,  0.6462655 , -1.1305623 ,
       0.15471256, -0.7179126 , -0.15460503, -0.32296193,  0.1598994 ,
       -0.5075027 , -0.7410554 ,  0.5247071 ,  0.0338988 ,  0.06148871,
       -0.26969084, -0.40099415,  0.30752096,  0.11141815, -0.5975455 ,
       0.9567183 ,  0.71071416, -0.3935786 , -0.60557264,  0.17251973,
       0.42103177,  0.02287591,  0.02722805, -0.9834109 , -0.35429567,
       0.23459306,  0.19908333,  0.42050263, -0.80189645,  0.11037309,
       -0.26703724,  0.9878612 ,  0.63543224, -0.02554742,  0.96865296,
       0.15772453,  0.29907194, -0.0074525 , -1.3398806 ,  0.17438231,
       -0.1138156 ,  0.4435544 , -0.30574077,  0.52437586,  0.2716861 ,
       -0.13374457,  0.034655 ,  0.8558606 ,  0.20205818, -0.17715032,
       0.5058484 ,  0.09029104,  0.16534522, -0.28038573,  1.0729178 ,
       1.0143198 , -0.3190712 , -0.8396994 ,  0.5295455 ,  0.6986663 ],
      dtype=float32)
```



# Preprocessing

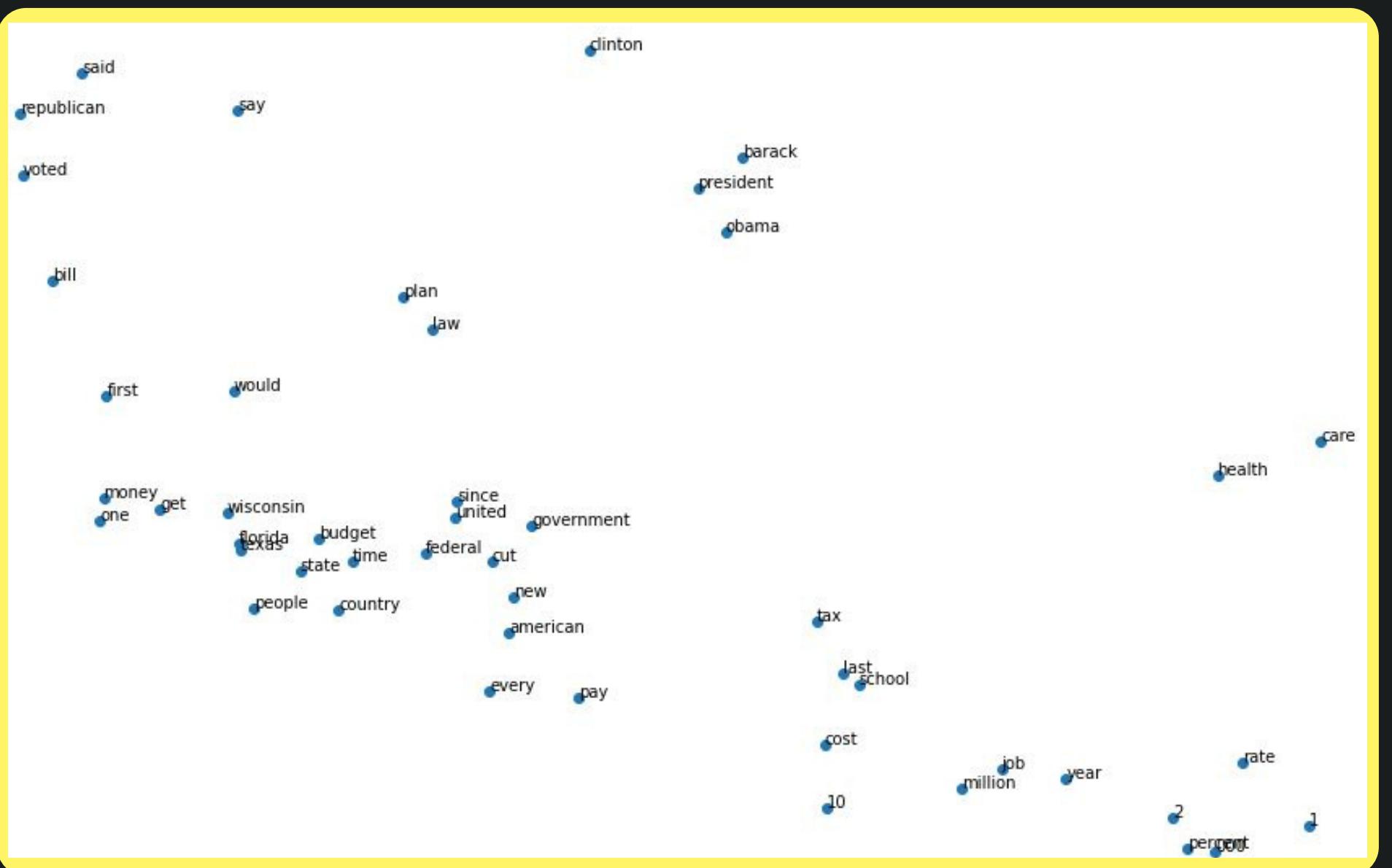
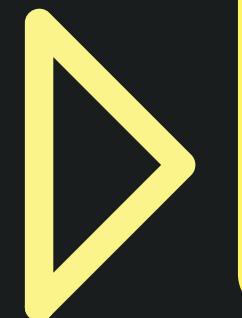
## Word Embedding

- Word2Vec



health

```
array([-0.95737183,  0.30383423,  0.36267897, -0.3145774 , -0.40760833,
       -0.9460593 ,  0.49187672,  1.4210457 , -0.30092633, -0.20819204,
       -0.770027 , -0.7723512 , -0.33388203, -0.1773534 ,  0.20409673,
       -0.5343583 , -0.39575577, -0.93514854, -0.06234565, -1.7523842 ,
       0.22604117,  0.65160227,  0.5052533 , -0.26991683, -0.3610879 ,
       -0.53496146, -0.32183996, -0.92692065,  0.07382824,  0.26401505,
       0.7777989 ,  0.04584725,  0.21644872, -0.6334908 ,  0.10794742,
       1.0837792 ,  0.05388981,  0.19897468,  0.6462655 , -1.1305623 ,
       0.15471256, -0.7179126 , -0.15460503, -0.32296193,  0.1598994 ,
       -0.5075027 , -0.7410554 ,  0.5247071 ,  0.0338988 ,  0.06148871,
       -0.26969084, -0.40099415,  0.30752096,  0.11141815, -0.5975455 ,
       0.9567183 ,  0.71071416, -0.3935786 , -0.60557264,  0.17251973,
       0.42103177,  0.02287591,  0.02722805, -0.9834109 , -0.35429567,
       0.23459306,  0.19908333,  0.42050263, -0.80189645,  0.11037309,
       -0.26703724,  0.9878612 ,  0.63543224, -0.02554742,  0.96865296,
       0.15772453,  0.29907194, -0.0074525 , -1.3398806 ,  0.17438231,
       -0.1138156 ,  0.4435544 , -0.30574077,  0.52437586,  0.2716861 ,
       -0.13374457,  0.034655 ,  0.8558606 ,  0.20205818, -0.17715032,
       0.5058484 ,  0.09029104,  0.16534522, -0.28038573,  1.0729178 ,
       1.0143198 , -0.3190712 , -0.8396994 ,  0.5295455 ,  0.6986663 ],
      dtype=float32)
```



# Preprocessing

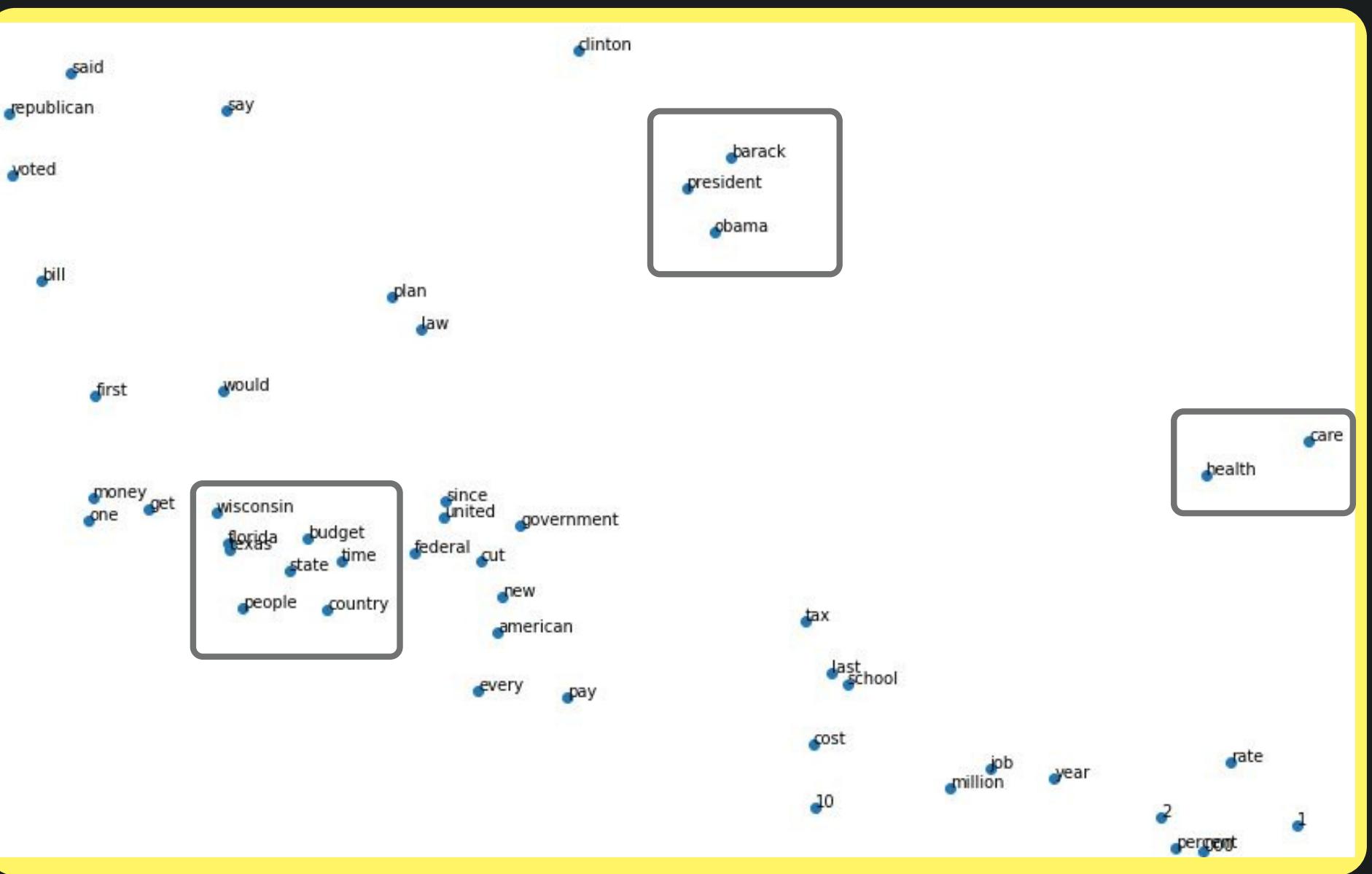
# Word Embedding

- Word2Vec

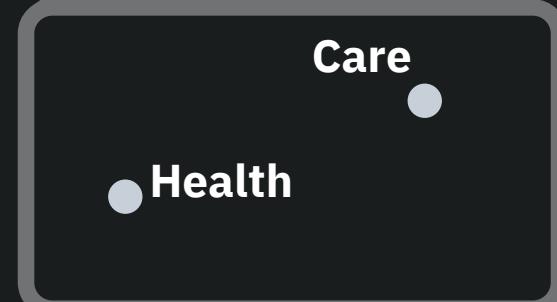
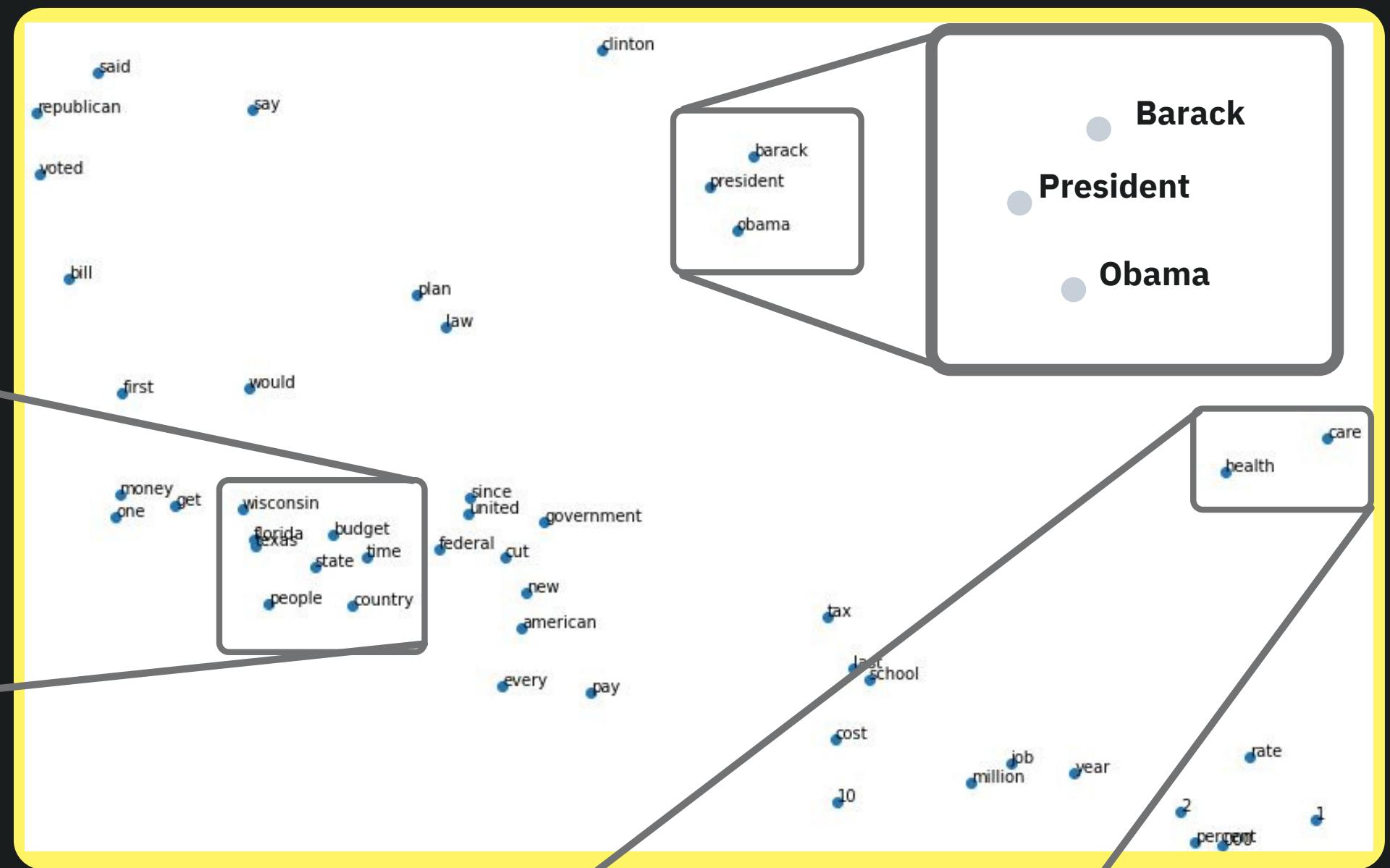
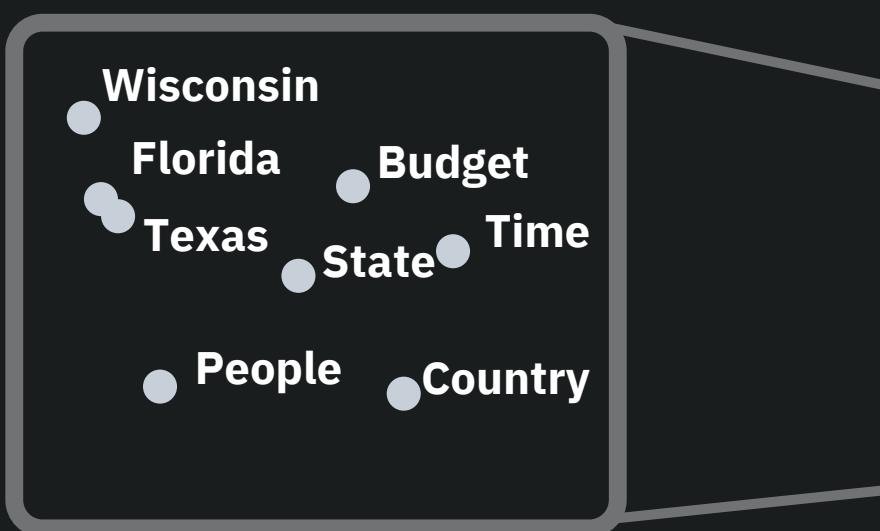
A yellow equilateral triangle with a black border, centered at the bottom of the page.

# health

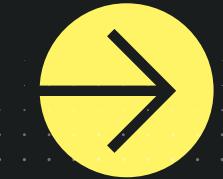
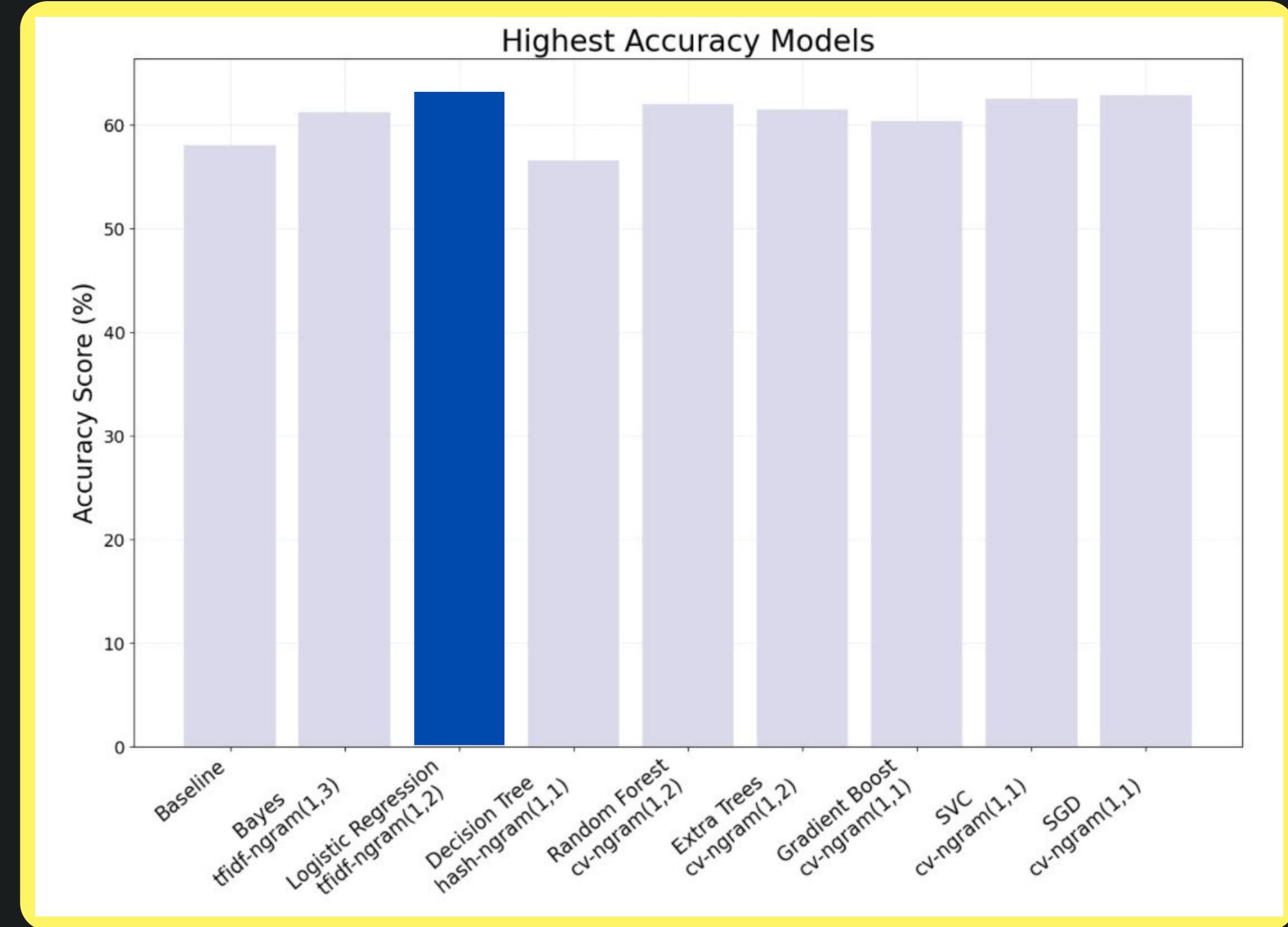
```
array([-0.95737183,  0.30383423,  0.36267897, -0.3145774 , -0.40760833,
       -0.9460593 ,  0.49187672,  1.4210457 , -0.30092633, -0.20819204,
       -0.770027 , -0.7723512 , -0.33388203, -0.1773534 ,  0.20409673,
      -0.5343583 , -0.39575577, -0.93514854, -0.06234565, -1.7523842 ,
      0.22604117,  0.65160227,  0.5052533 , -0.26991683, -0.3610879 ,
     -0.53496146, -0.32183996, -0.92692065,  0.07382824,  0.26401505,
      0.7777989 ,  0.04584725,  0.21644872, -0.6334908 ,  0.10794742,
      1.0837792 ,  0.05388981,  0.19897468,  0.6462655 , -1.1305623 ,
      0.15471256, -0.7179126 , -0.15460503, -0.32296193,  0.1598994 ,
     -0.5075027 , -0.7410554 ,  0.5247071 ,  0.0338988 ,  0.06148871,
     -0.26969084, -0.40099415,  0.30752096,  0.11141815, -0.5975455 ,
      0.9567183 ,  0.71071416, -0.3935786 , -0.60557264,  0.17251973,
      0.42103177,  0.02287591,  0.02722805, -0.9834109 , -0.35429567,
      0.23459306,  0.19908333,  0.42050263, -0.80189645,  0.11037309,
     -0.26703724,  0.9878612 ,  0.63543224, -0.02554742,  0.96865296,
      0.15772453,  0.29907194, -0.0074525 , -1.3398806 ,  0.17438231,
     -0.1138156 ,  0.4435544 , -0.30574077,  0.52437586,  0.2716861 ,
     -0.13374457,  0.034655 ,  0.8558606 ,  0.20205818, -0.17715032,
      0.5058484 ,  0.09029104,  0.16534522, -0.28038573,  1.0729178 ,
      1.0143198 , -0.3190712 , -0.8396994 ,  0.5295455 ,  0.6986663 ],
      dtype=float32)
```



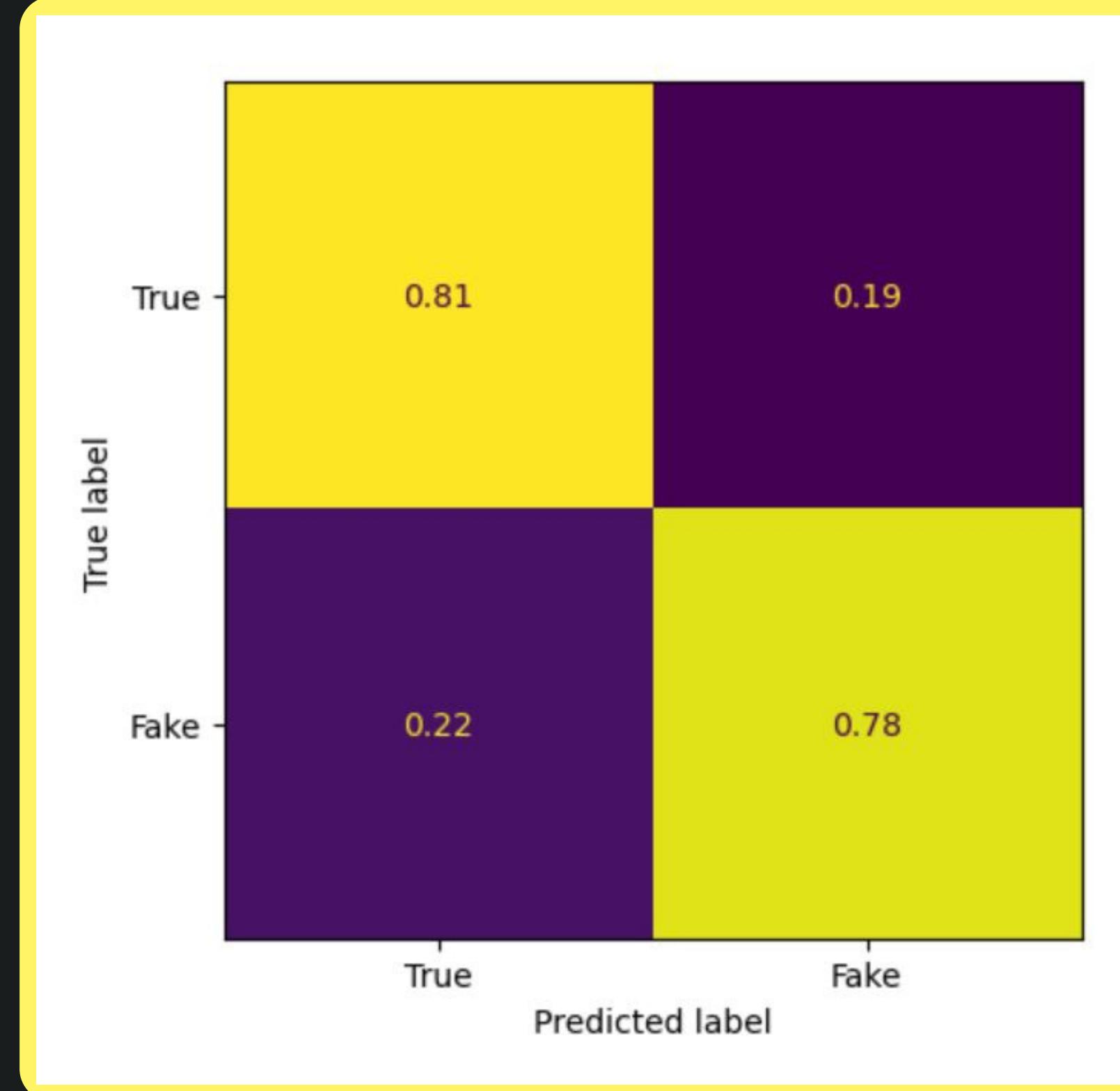
# Preprocessing



# Models Performances



# Best Model Confusion Matrix



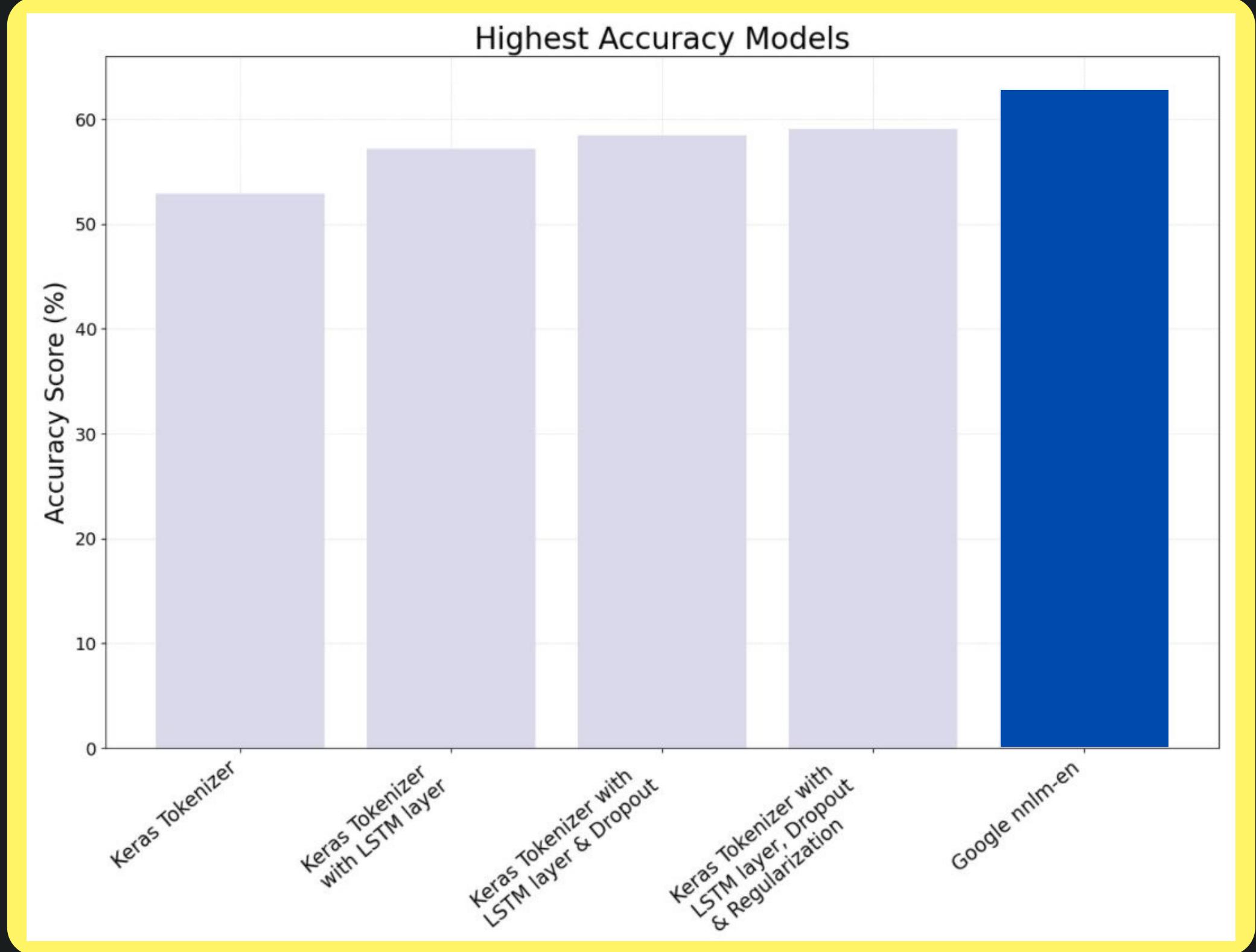
Accuracy

~63%

Increase from  
Baseline

~5%

# Neural Network Models Performances



# Next Steps

## Different Languages

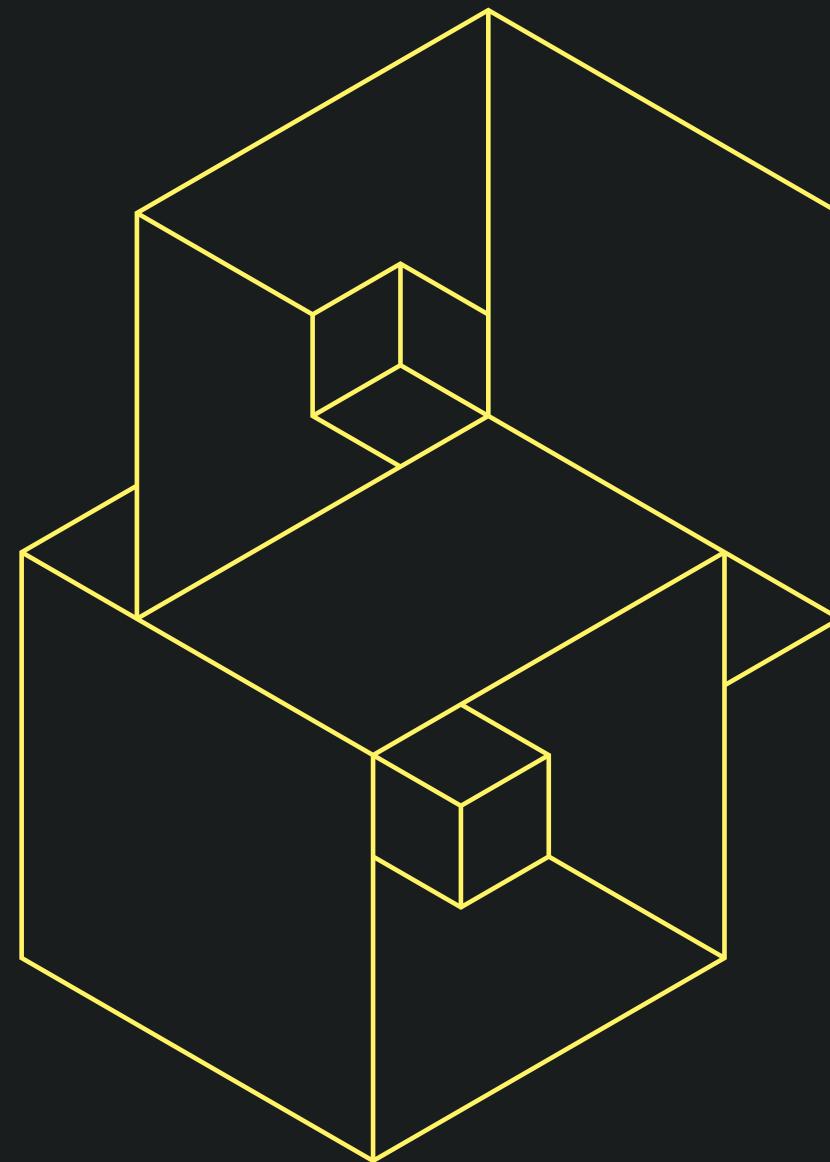
- The model can be adapted to perform in different languages if there is a pre-trained model with that language vocabulary embeded

## Use Metadata

- The person making the statement, their political affiliation and other characteristics may affect how the model operates. If we have that information for new statements, it could give us a more accurate classification



# Thank you



**Marcelo Scatena**



[marcelo.odd@gmail.com](mailto:marcelo.odd@gmail.com)



[github.com/moscatena](https://github.com/moscatena)

