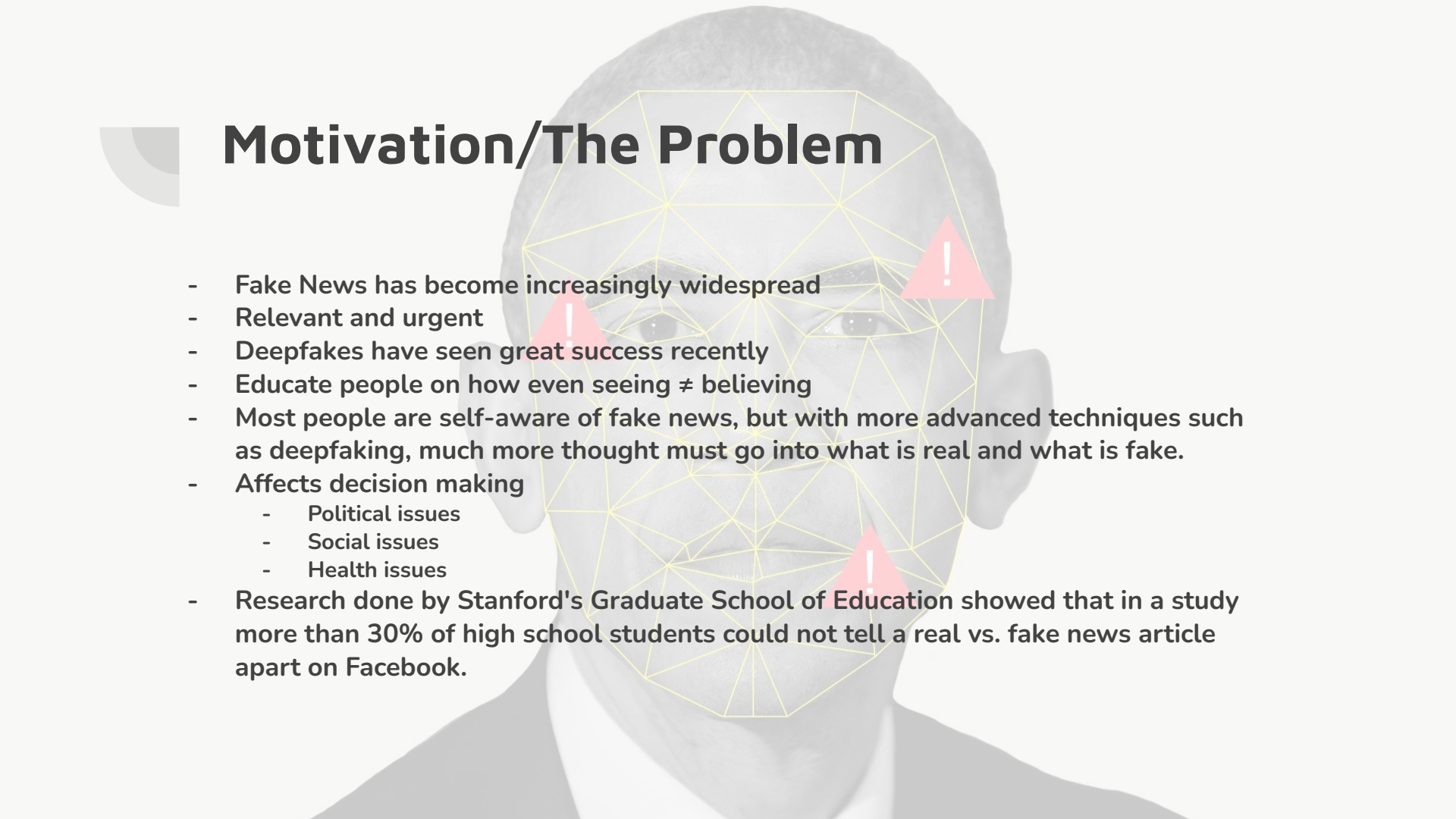
Fake News Detection

Justin Ecarma, Tara Adusumilli, Kiron Das,
Isaac Jiang, Aaron Qi

Motivation/The Problem



- Fake News has become increasingly widespread
- Relevant and urgent
- Deepfakes have seen great success recently
- Educate people on how even seeing ≠ believing
- Most people are self-aware of fake news, but with more advanced techniques such as deepfaking, much more thought must go into what is real and what is fake.
- Affects decision making
    - Political issues
    - Social issues
    - Health issues
- Research done by Stanford's Graduate School of Education showed that in a study more than 30% of high school students could not tell a real vs. fake news article apart on Facebook.

Real or Fake (type in chat)

https://www.theguardian.com/newswise/2021/feb/04/fake-or-real-headlines-quiz-newswise-2021

1. Nasa is installing internet on the moon

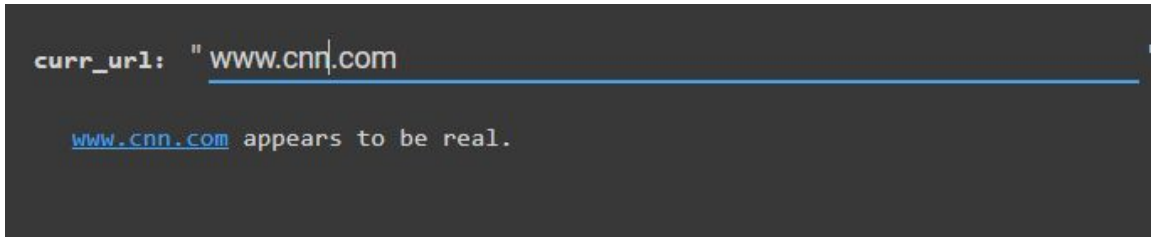
7. Spinach is taught how to send emails

Data & Training

- Train & Test Data
- Methods of training:
  - URL extensions (.com, .net, ect...)
  - Common Keyword abundance
  - Bag of words
  - Word vectors
- Accuracy around 85% using via a logistic regression model
  - Combining keywords, bow, and word vectors into one variable
  - This is by using test data vs what it learned from the training data

Architecture

- Import libraries and Data
- Train Data via the 4 methods
- User inputs website url
- Prediction returns a 0 or 1, 0 being real and 1 being fake



Results

- Some of the weights our model learned for various words and domains:

```
[('associated press keyword', -1.6427300931550461),
 ('ap keyword', -1.4539143344911445),
 ('finance keyword', -1.4031648323999648),
 ('2021 keyword', -1.0861853011684823),
 ('2020 keyword', -0.9161038375434299),
 ('sports keyword', -0.885184218006038),
```

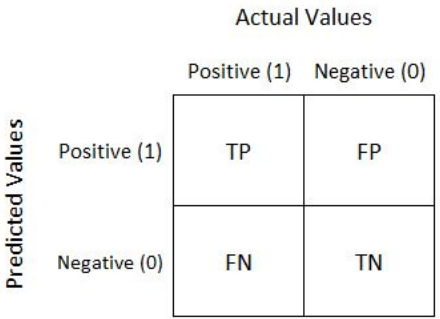
```
('obama keyword', 1.1208605833034082),
('hillary keyword', 1.1955130376504253),
('/ keyword', 1.209934689581746),
('.co domain', 1.4736510789136839),
('.info domain', 1.5460482344246504),
('.net domain', 2.247389068965282),
('.news domain', 2.618041563620683)]
```

```
('nationalism keyword', -0.09014720785383504),
('.com domain', -0.07033277615022722),
('middle east keyword', -0.038380980902270004),
('america keyword', -0.003360083140738654),
('facism keyword', 0.0),
('fear keyword', 0.036688153997012384),
('biden keyword', 0.09805442255381207),
```

Results (cont)

Confusion Matrices:

- Demonstrate the distribution of True Positives, False Positives, False Negatives, and True Negatives.
- 0 is real, 1 is fake




```
Confusion matrix:
[[ 77  91]
 [ 13 128]]
```

Conclusion

- Fake news and misleading information interfere with the user's ability to discern useful information from the Internet when news becomes critical for decision making.
  - ➢ The issue of fake news has become more widespread in this day and age.
- Our project utilizes machine learning models and NLP with news articles as training data, more specifically the domain extension and keywords.
  - ➢ Our model effectively identifies fake news with ~85% accuracy.
- The production of fake news will only increase as time goes on, so it is up readers to differentiate between real/fake news and our tool helps do exactly that.