PRESENTATION ON

FAKE NEWS DETECTION

Under the Guidance of

Mr. Abhinav Gupta, Mr. Ravish Kumar Dubey & Ms. Jayati Bhardwaj

What is Fake News Detection?

Fake News Detection is a tool or platform that detect information content that is false, misleading or whose source cannot be verified. This content may be generated to intentionally damage reputations, deceive, or to gain attention.

Misinformation presents a huge challenge in online society. As a result, there have been many attempts to identify and classify misinformation. Specifically, in social networking sites, blogs, as well as online newspapers.

Which libraries we've used?

- **NumPy**
- **Pandas**
- **Matplotlib**
- **Seaborn**
- **Sklearn**

- **NumPy:** NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays.
- **Pandas:** Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.
- **Matplotlib:** Matplotlib is one of the most popular Python packages used for data visualization.
- **Seaborn:** Seaborn is a library mostly used for statistical plotting in Python.
- **Sklearn:** Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

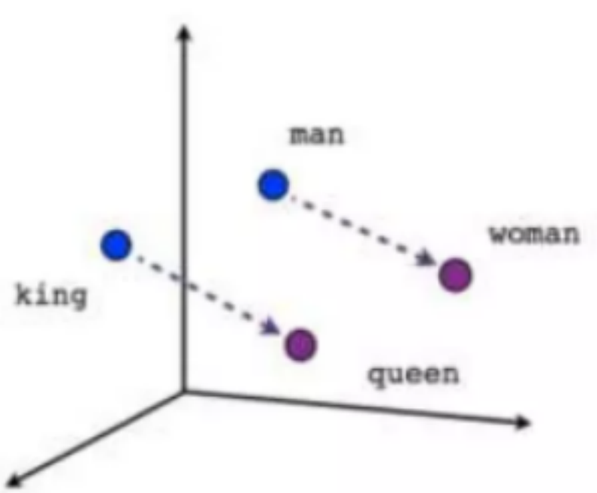
Fake News Is A Real Problem

Facebook engagement of the top five fake election stories*

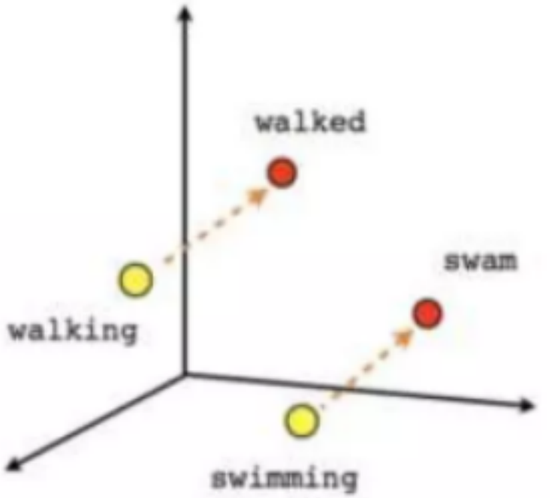
How we set up our problem and data

- We input the **domain names, HTML codes, and labels** (0- true, 1-fake) of 2002 news websites to **train** the model.
- Then we used **309** examples to **test** it. The output were **labels of either 0 or 1 i.e. real or fake.**

Under the broad domain on ‘Natural Language Processing’, we used different approaches like **keyword search, Bag of Words, and GloVe.**



Male-Female



Verb tense

‘Vectors’ of words

Our best model

A combination of bag-of-words, word-2-vector, and the feature description model was found to be the one which showed us maximum accuracy.

- BAG OF WORDS- looks at the count of each word.
- WORD-2-VEC-finds out the actual meaning of the world
- FEATURE DESCRIPTION- looking at the ‘url’, and the ‘html’ to infer the difference fake and real news.

```
test_X = []
for url, html, label in test_data:
  curr_X = np.array(featurize_data_pair(url, html))
  test_X.append(curr_X[0])

test_X = np.array(test_X)

test_y = [label for url, html, label in test_data]

print('Done loading test data...')

test_y_pred = model.predict(test_X)

print('Test accuracy', accuracy_score(test_y, test_y_pred))

print('Confusion matrix:')
print(confusion_matrix(test_y, test_y_pred))

prf = precision_recall_fscore_support(test_y, test_y_pred)

print('Precision:', prf[0][1])
print('Recall:', prf[1][1])
print('F-Score:', prf[2][1])

### END CODE HERE ###
```

```
100%|██████████| 399809/400000 [01:00<00:00, 8726.68it/s]Loading test data.
Done loading test data...
Test accuracy 0.8739837398373984
Confusion matrix:
[[104  30]
 [  1 111]]
Precision: 0.7872340425531915
Recall: 0.9910714285714286
F-Score: 0.8774703557312252
```