



Detecting Fake News

Machine Learning Model

May I Help
You?

The Age of Disinformation



The Source

The Problem



The Solution



Fake News
Analyzer

Fake New Analyzer

Our Premise is to apply Machine Learning to News Articles in order Detect “Fake News!”



Passive Aggressive Classifier

The Passive-Aggressive algorithms are a family of Machine Learning algorithms that fall under online learning, where you train the model incrementally and continuously. It will self-adjust the model based on whether predictions were correct (passive) or not (aggressive).

- **Passive:** If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
- **Aggressive:** If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it.

Passive Aggressive Classifier

Advantage: Small memory footprint as the model sequentially process data as received.

Disadvantage: There is no “big picture” of processed data.

- End result is affected by the order of data presentation.
- Model is as good as the quality of training data.

TfidfVectorizer

Used to convert text to numerical representation
to fit Passive-Aggressive Classifier.

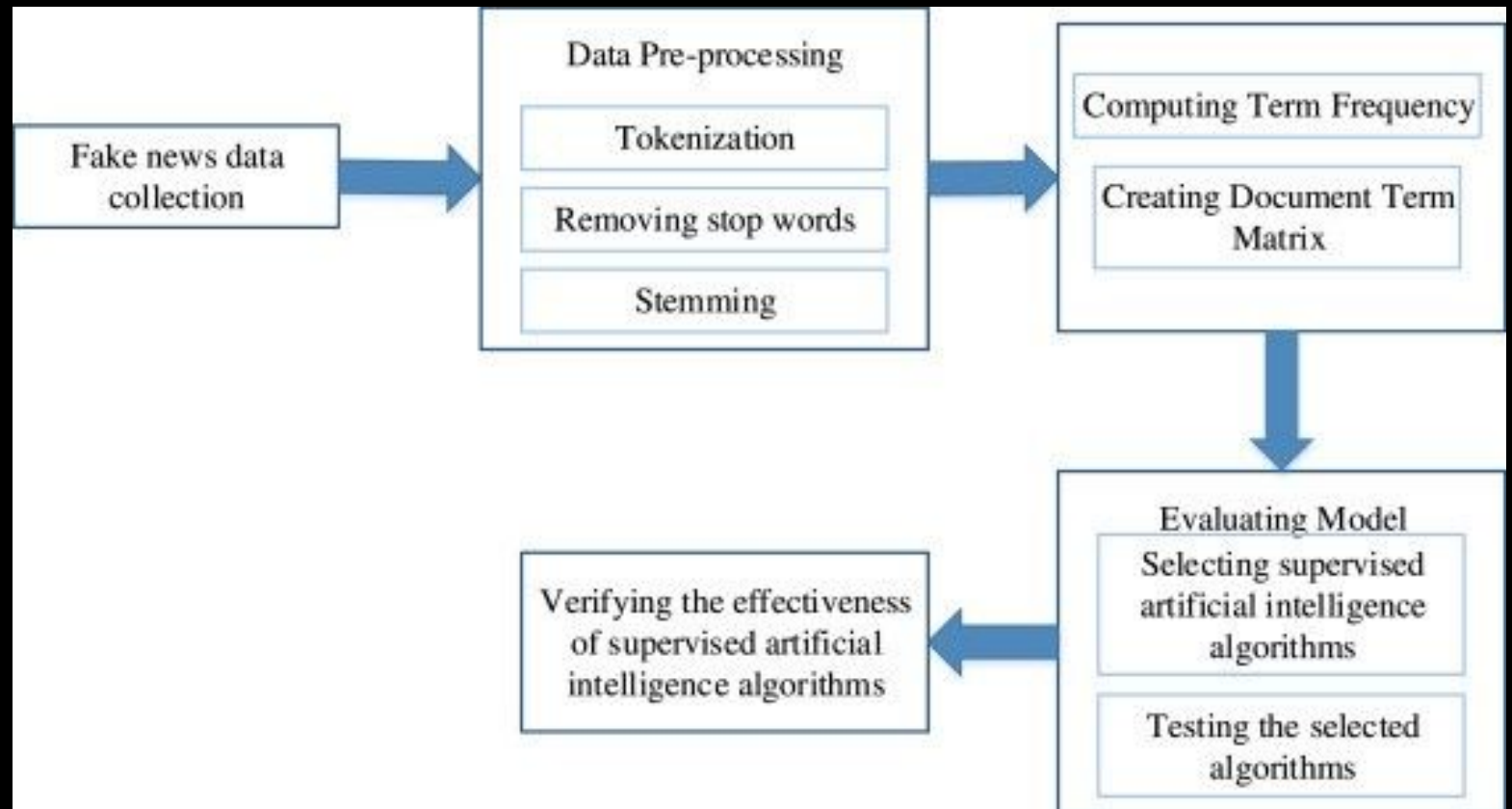
Term-Frequency (tf)

$tf(w) = \text{doc.count}(w) / \text{total words in the doc}$

Inverse Document Frequency (idf)

$idf = \log(\text{total_num_of_docs} / \text{number_of_docs_containing_word_w})$

$td-idf(w) = tf(w) * idf(w)$



Implementation – *Jupyter Notebook*

<https://jupyter.org/>

❖ An interactive, open-source web app that allows you to create documents containing live code, equations and visualizations

Took 4 Step Approach

❖ Data Sourcing

- Find “Evaluated” Data (baseline) *Key Thought*

❖ Preparation

- Clean & Scrub with pandas and numby, then Split the Data 80/20

❖ Train Model

- Train Model with 80 Split

❖ Accuracy Testing

- Test the Model with 20 Split

Training Data Sample

- Dataset of over 6,000 articles used with 0.2 test_size ratio (80-20 split between training and testing).
- Accuracy of 92.74%

```
# Note the last column denoting whether articles are fake or real
pd.set_option('display.max_colwidth', 500)
data.head()
```

✓ 0.3s

Python

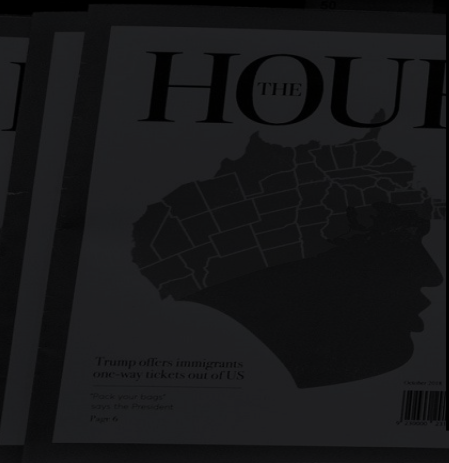
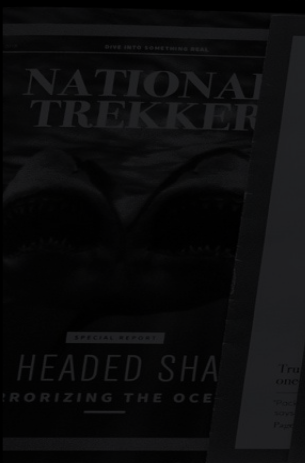
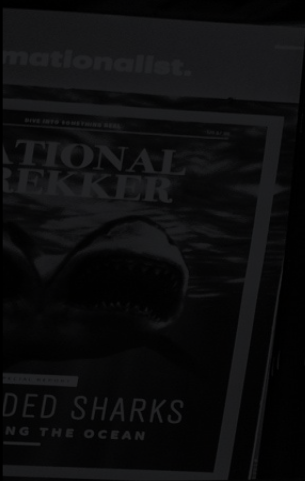
	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a New York writer focusing on radical Islam. \n\nIn the final stretch of the election, Hillary Rodham Clinton has gone to war with the FBI. \nThe word "unprecedented" has been thrown around so often this election that it ought to be retired. But it's still unprecedented for the nominee of a major political party to go war with the FBI. \nBut that's exactly what Hillary and her people have done. Coma patients just waking ...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Political Suicide At A Trump Rally (VIDEO)	Google Pinterest Digg Linkedin Reddit Stumbleupon Print Delicious Pocket Tumblr \n\nThere are two fundamental truths in this world: Paul Ryan desperately wants to be president. And Paul Ryan will never be president. Today proved it. \n\nIn a particularly staggering example of political cowardice, Paul Ryan re-re-re-reversed course and announced that he was back on the Trump Train after all. This was an aboutface from where he was a few weeks ago. He had previously declared he would not be suppor...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sunday's unity march against terrorism. \n\nKerry said he expects to arrive in Paris Thursday evening, as he heads home after a week abroad. He said he will fly to France at the conclusion of a series of meetings scheduled for Thursday in Sofia, Bulgaria. He plans to meet the next day with Foreign Minister Laurent Fabius and President Francois Ho...	REAL

Server Side

- Flask driven web app
- Pre-Processing
 - Scikit-Learn for preparing data for use with Machine Learning (ML) model
 - Pandas for Series creation
- API for handling requests for analysis (JSON)

Front End

- Single Page Application (SPA)
- Leverages Bootstrap / Custom CSS / Font-Awesome
- JavaScript (Client-Side)





Detecting Fake News

Machine Learning Model