Fake News Detection Using NLP

Join me in exploring how natural language processing can be used to identify fake news, from data source to model evaluation.





The New York Time

Diver Work Times

TW YORK, WEDNESDAY, NOVEMBER 9.

CHESTER BY

P WINS IN LAND

EDIATE ACTION ON EXECUT

E: "EXHAUSTED" Ly. Says bill

conditional is non-likewood to a size, the number's would be the territor with discours and party sy's creeding distinct, another to

Interiors and otheral lapses do
most social compagit was a
sowed domains creating. Visite
most should be domain. Visite
most fact that the compagit
classed speech domain as Agricultural
speech domain. DAS
factors DNA creation in favor
Ms Classes in skelder podelices. Even then controvers
bits Classes "D- America.

at a Spacell berindle landshoot freezishers olsoy. Thronty's lands pass and the venture lands of the state of

med the featon to accommendable actions in Very Williams counter for the long-grain, Himselopes of grant mandot, and the 1 pant, You say blamous, Ny to hocking organic rises before lambly mandoosi.

President elect Donald J. Trump joined supporters at a raily in Washingto

SCAVINO EMERGES AS POLITICAL POWER

BU ICATHOY BANGEBEE CAPE

A reliative unknown in Weshington political circles prior to his appointment as wrater as Transc. Dented Septime. It, has alconstituted to factor in recent modes, coming anxiog as

Data Source

Social Media

Twitter, Facebook, and other social media platforms are popular sources of news and information.

News Outlets

Traditional news organizations often have online sites, which provides a wealth of data.

Web Scraping

Web scraping tools can extract data from any publicly available source online.

Data Preprocessing



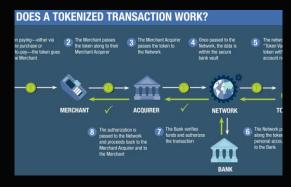
Data Cleaning

Remove any irrelevant information, such as HTML tags and URLs.



Stemming & Lemmatizing

Reduce words to their root stem or base form so that they can be analyzed more accurately.



Data Tokenization

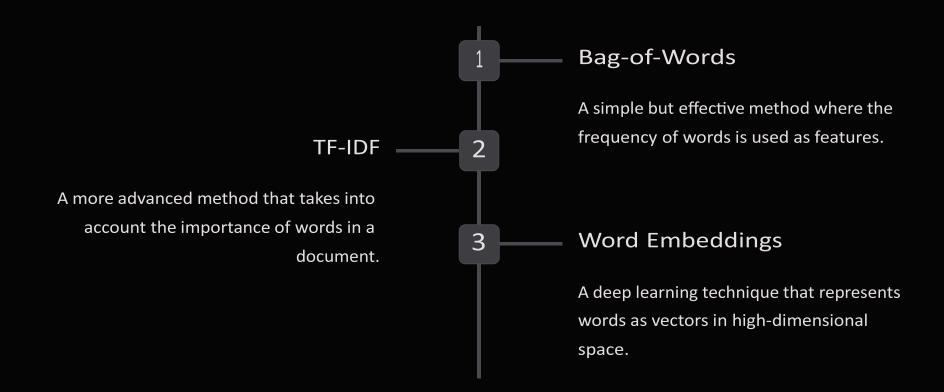
Split up the data into words or phrases so that it can be easily analyzed.



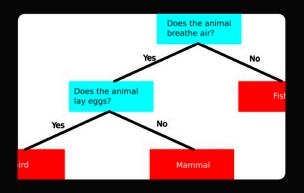
Stopword Removal

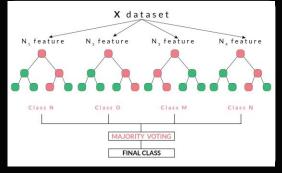
Filter out common words that don't provide useful information.

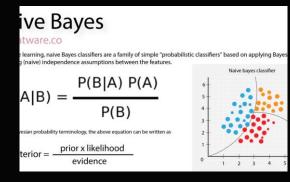
Feature Extraction



Model Selection







Decision Tree

Creates a tree-like model of decisions and their possible consequences.

Random Forest

Creates multiple decision trees to improve accuracy and reduce overfitting.

Naive Bayes

A probabilistic algorithm that makes predictions based on a set of probabilities.

Model Training

1 Split Data

Divide the dataset into a training set and a testing set. The training set teaches the model and the testing set evaluates its accuracy.

2 Train the Model

Feed the training data into the selected algorithm and adjust the parameters to optimize performance.

3 Repeat

Train and test the model multiple times, adjusting hyperparameters each time until optimal accuracy is achieved.



Model Evaluation

Accuracy

How well the model predicted the correct class.

Recall

How often the model correctly identified a class when it was actually that class.

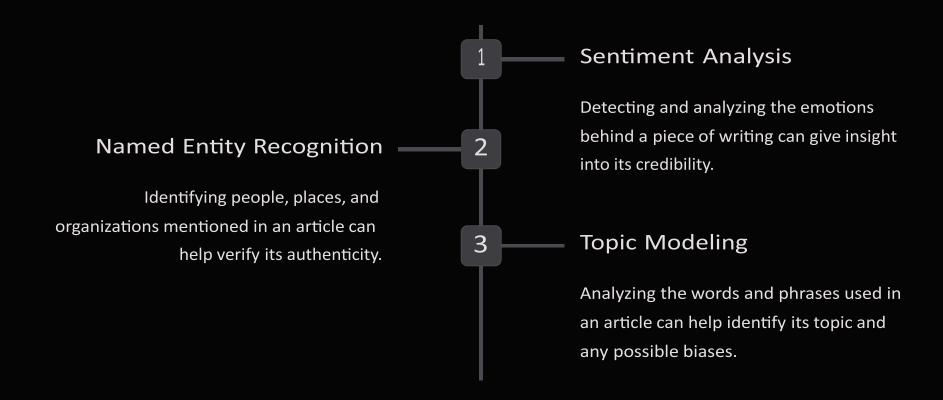
Precision

How often the model predicted the correct class when it actually was that class.

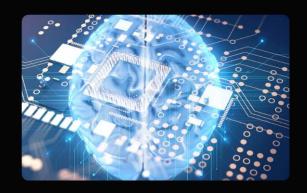
F1 Score

A combination of recall and precision that takes into account false positive and false negative rates.

Application of NLP for Fake News Detection



The Future of Fake News Detection



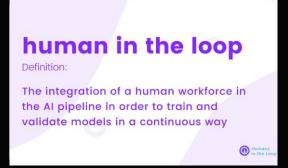
Deep Learning

Using neural networks with many layers to detect complex patterns and relationships in data.



Explainable Al

Developing models that provide transparent explanations for their decisions.



Human-in-the-Loop

Combining the strengths of Al and human judgement to create more accurate models.