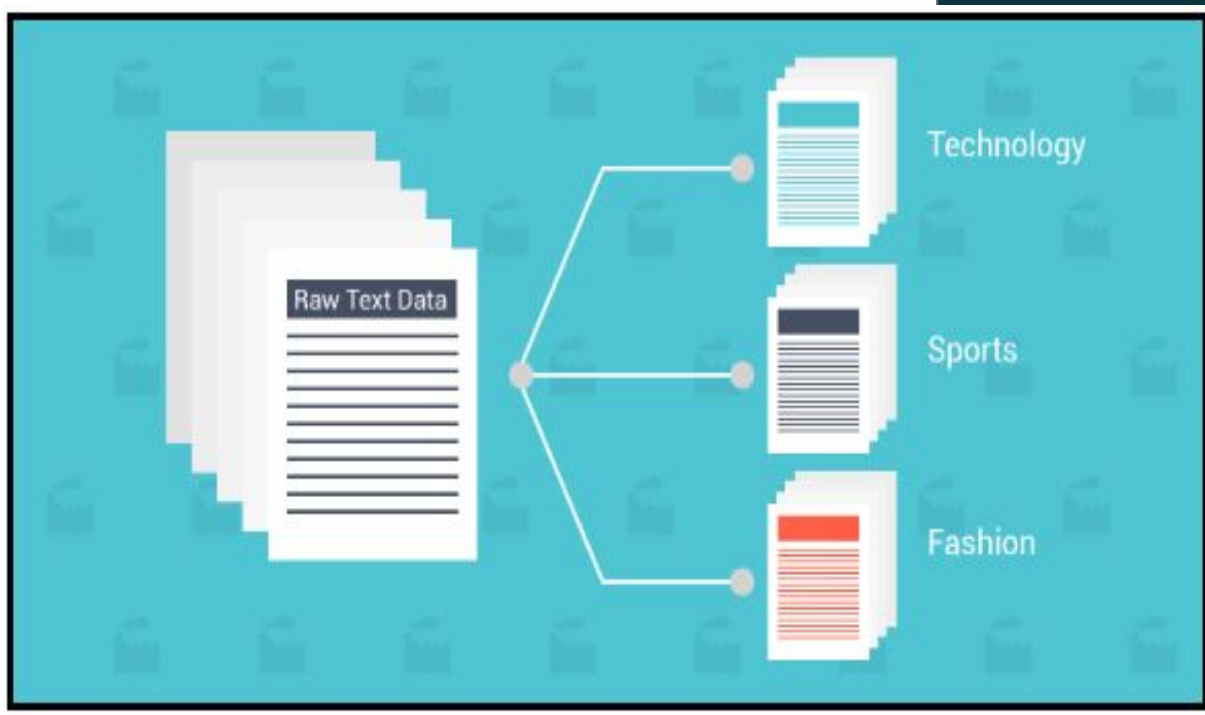




Using **NLPs**

By **Bram Tungala**



The goal of text classification is to classify text documents into one or more predefined categories.

- Categorize news articles
- Sentiment analysis
- Spam detection
- Credit card fraud detection
- Auto tagging customer request

Natural Language Processing (NLP) and Supervised Machine Learning are widely used in ***Text Classification***.

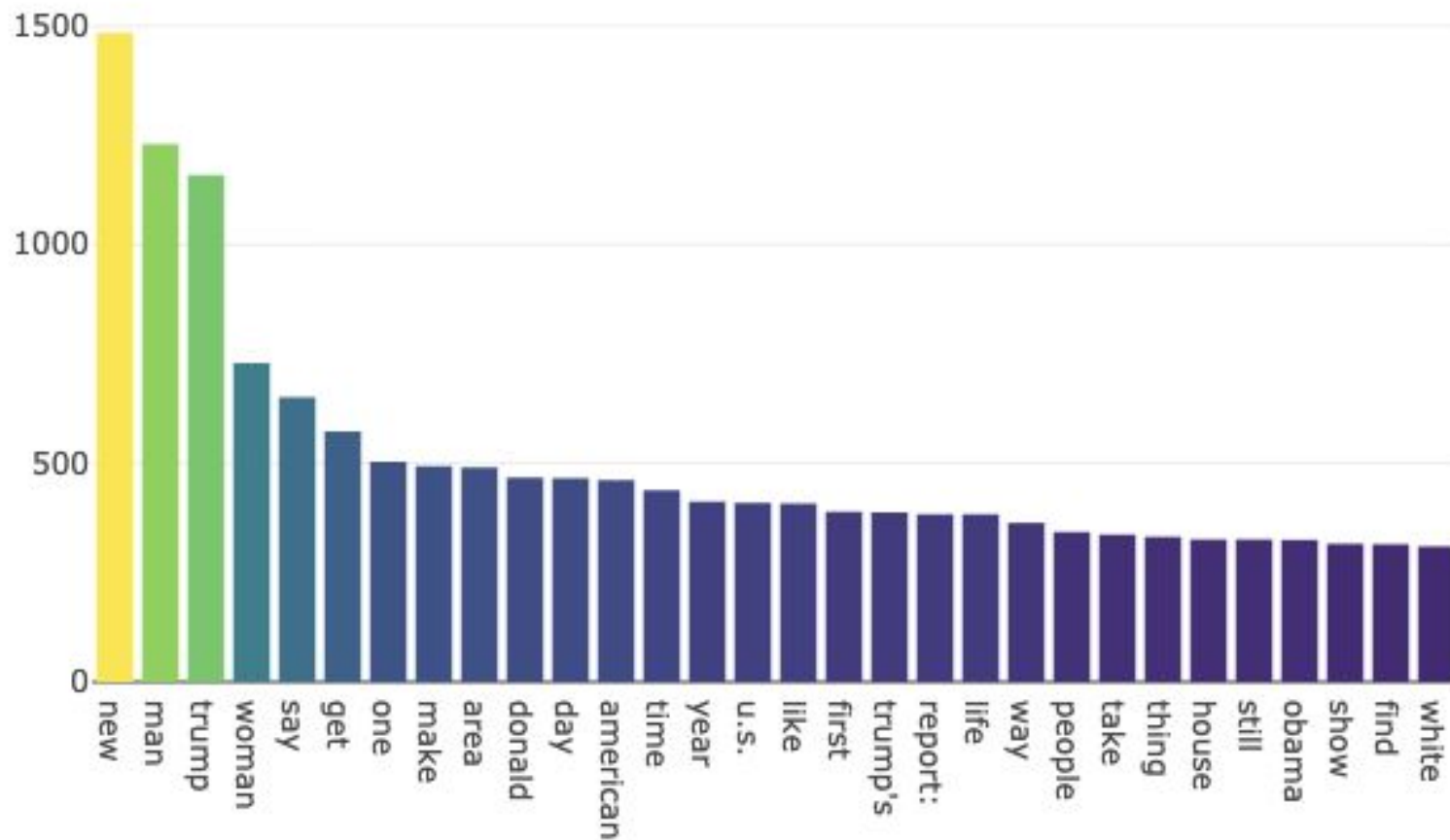
**Can we use machine learning to
detect fake news from real news
using News Headlines?**

Data

- Source
<https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection>
- 26k+ records
- Two types of news (pre-labeled):
 - Fake: TheOnion
 - Real: Huffpost

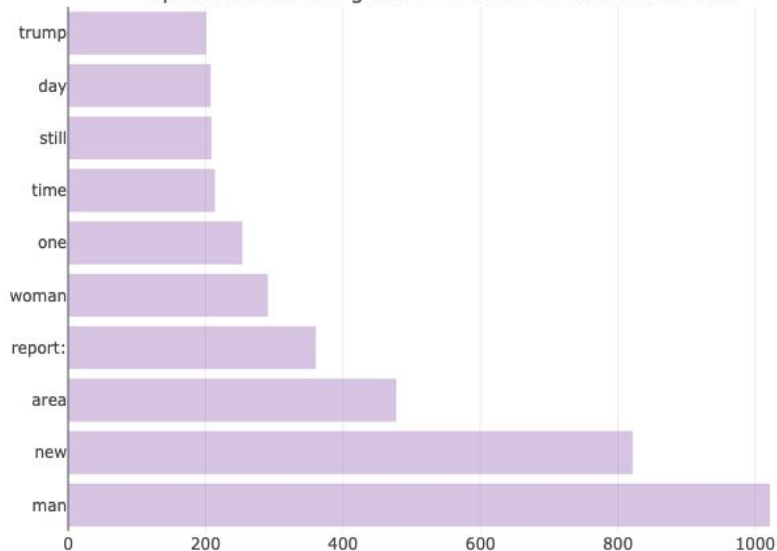
[illegible]

Word count - Top 30 words

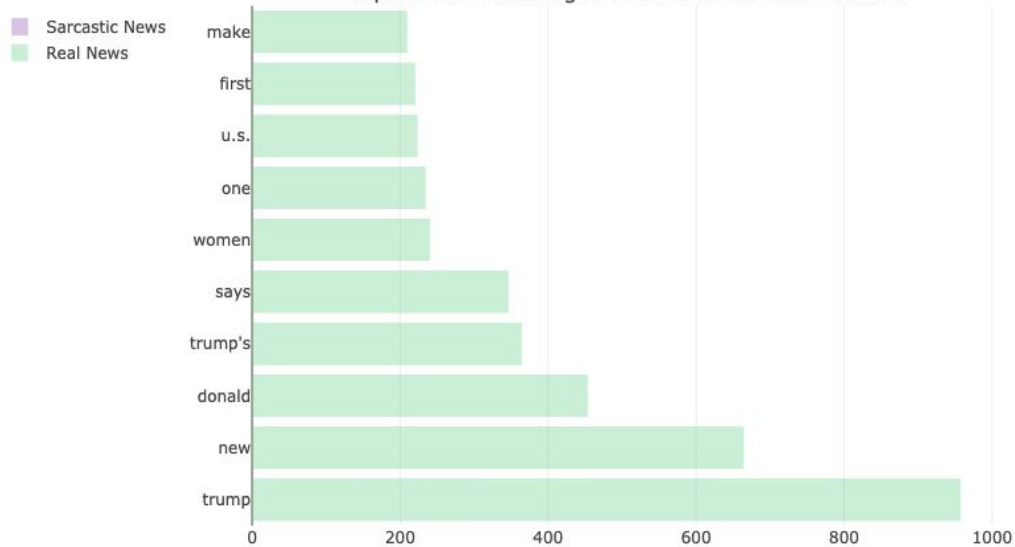


Top 10 words

Top 10 Most occurring words in Sarcastic News Headlines



Top 10 Most occurring words in Real News Headlines



Model 1: RNN

Accuracy: 81.3%

Model 2: Pre-trained RNN

Accuracy: 81.4%



Model 3: Pre-trained RNN + optimizer = RMSprop

Accuracy: 80%

Future Work

Tweak model parameters to see if we can improve accuracy and predictability.

- Use more pre-trained text data in the embeddings (GloVe)
- Use different tuning methods

A word cloud featuring the phrase "Thank You" in numerous languages and colors. The central, largest text is "thank you" in red. Other prominent words include "gracias" in green, "danke" in blue, "merci" in orange, and "teşekkür ederim" in pink. Smaller words in various colors include "spas", "ngiyabonga", "شكراً جزيلاً", "rahmat", "спасибо", "Баярлалаа", "faafetai lava", "nanni", "nandri", "kiitos", "dankie", "dhanyavad", "mauriuru", "koszonóm", "hvala", "gracie", "dziękuję", "sagolun", "sukriya", "kop khun krap", "go raibh maith agat", "arigatō", "takk", "dakujem", "trugarez", "merci", "shukriya", "merce", "mercsi", "xiexie", "감사합니다", "ευχαριστώ", "diolch", "dhanyavadagalu", "tanemirt", "rahmet", "najis tuke", "kam sah hamnida", "dijf madloba", "mesi", "sobodi", "dekuji", "chnorakaloutioun", "gratias ago", "gracies", "sulpáy", "akun", "dankon", "aciú", "misaotra", "matondo", "paldies", "grazzi", "mahalo", "tapadh leat", "xhala", "asante", "manana", "obrigada", "murakoze", "tenki", "chokrame", "mamnun", "djakую", "mochchakkeram", "bedankt", "enkosi", "bayarlalaa", "nandri", "nanni", "kiitos", "dankie", "dhanyavad", "mauriuru", "koszonóm", "hvala", "gracie", "dziękuję", "sagolun", "sukriya", "kop khun krap", "go raibh maith agat", "arigatō", "takk", "dakujem", "trugarez", "merci", "shukriya", "merce", "mercsi", "xiexie", "감사합니다", "ευχαριστώ", "diolch", "dhanyavadagalu", "tanemirt", "rahmet", "najis tuke", "kam sah hamnida", "dijf madloba", "mesi", "sobodi", "dekuji", "chnorakaloutioun", "gratias ago", "gracies", "sulpáy", "akun", "dankon", "aciú", "misaotra", "matondo", "paldies", "grazzi", "mahalo", "tapadh leat", "xhala", "asante", "manana", "obrigada", "murakoze", "tenki", "chokrame", "mamnun", "djakую", "mochchakkeram".