REAL VS. FAKE NEWS

Kevin Gross

CNN

Agenda

Real vs. Fake News

Overview

Business & Data Understanding

Modeling

Evaluation

Recommendation & Next Steps

CNN

Overview

Multinational 24-Hour Cable News Channel

Real vs. Fake News

"To Inform, Engage, and Empower the World"

Reputation at stake
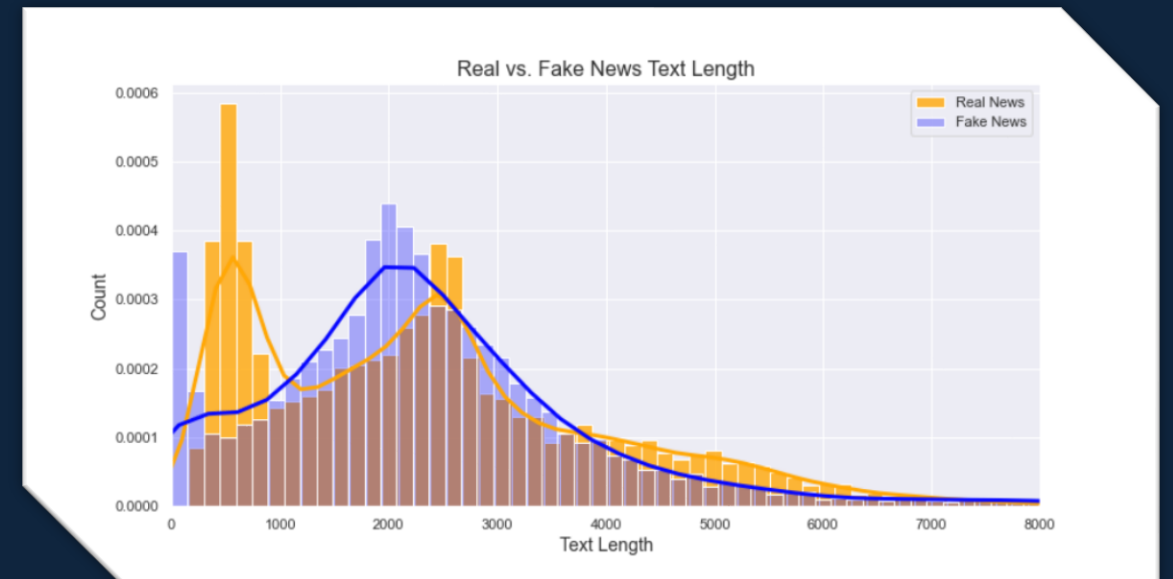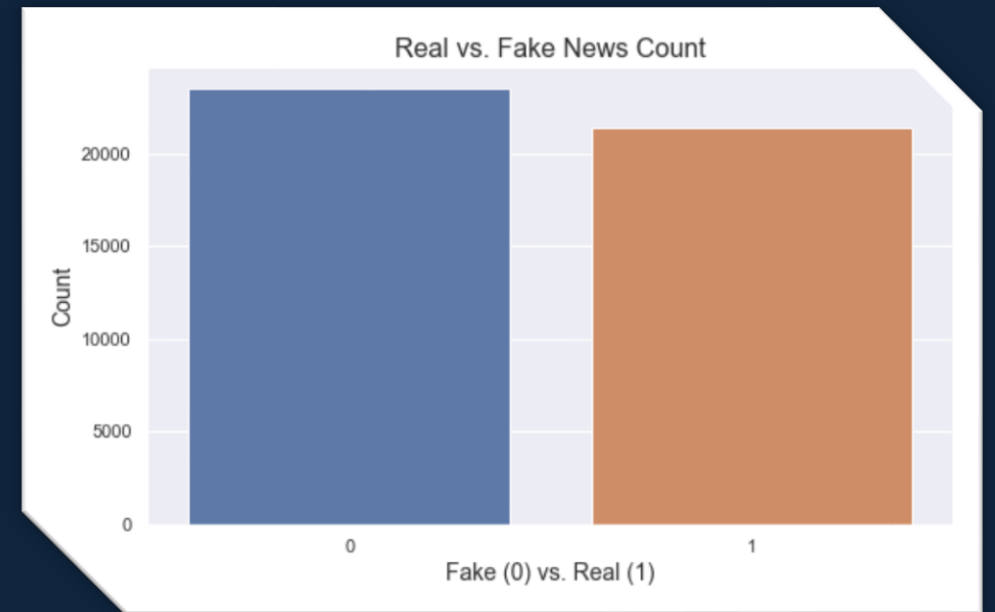
CNN

Business Understanding

Information overload

Fact, half-truth, or lies?

Digital platforms ripple effect

Real vs. Fake News

CNN

DATA
UNDERSTANDING

Real vs. Fake News Count

Real vs. Fake News Text Length
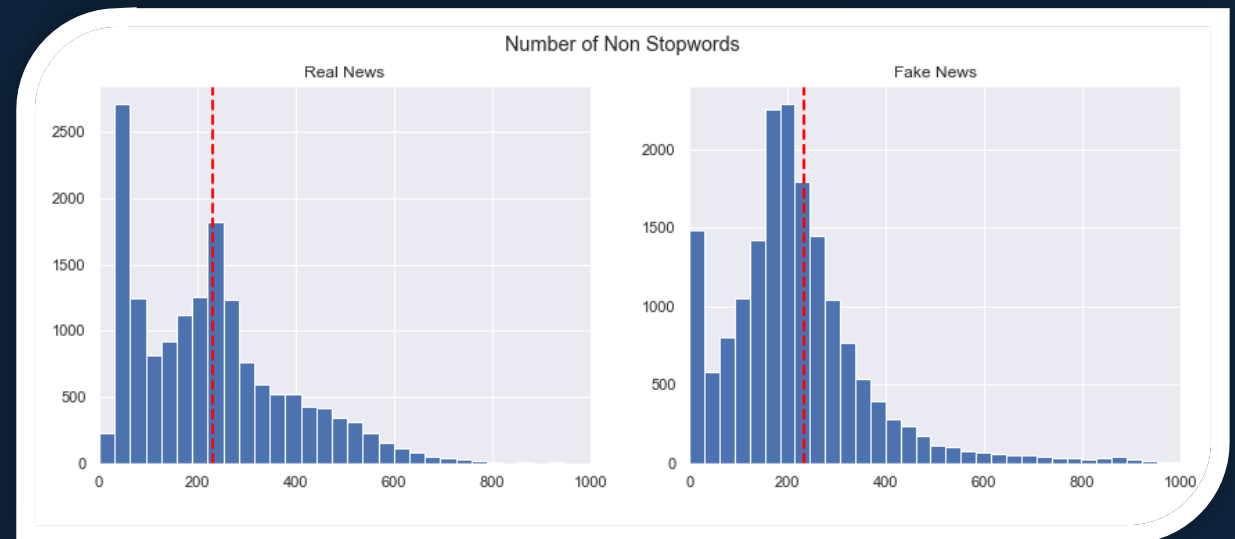
CNN

Modeling

Standardizing Case

Removing Punctuation

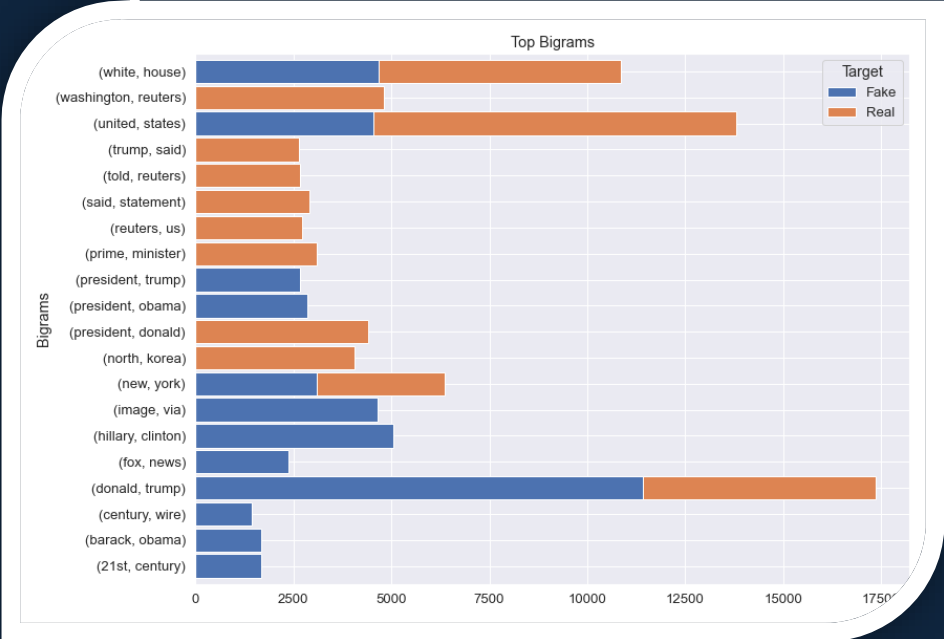Removing Stopwords

Stemming and Lemmatizing

Engineering Features

Real vs. Fake News

CNN

Engineered Features

Number of Sentences in the Article

Number of Non Stopwords
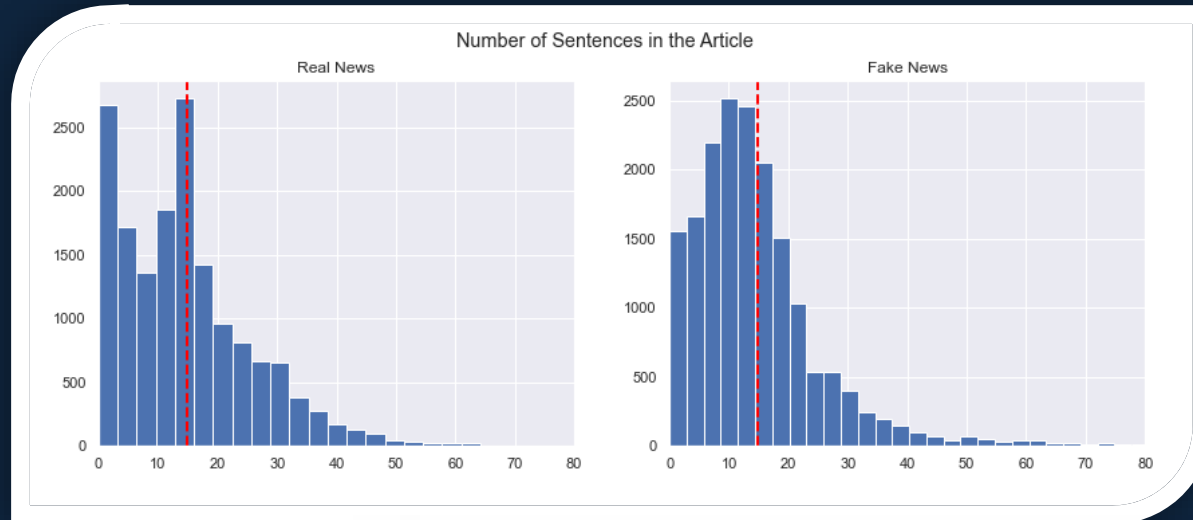
Engineered Features
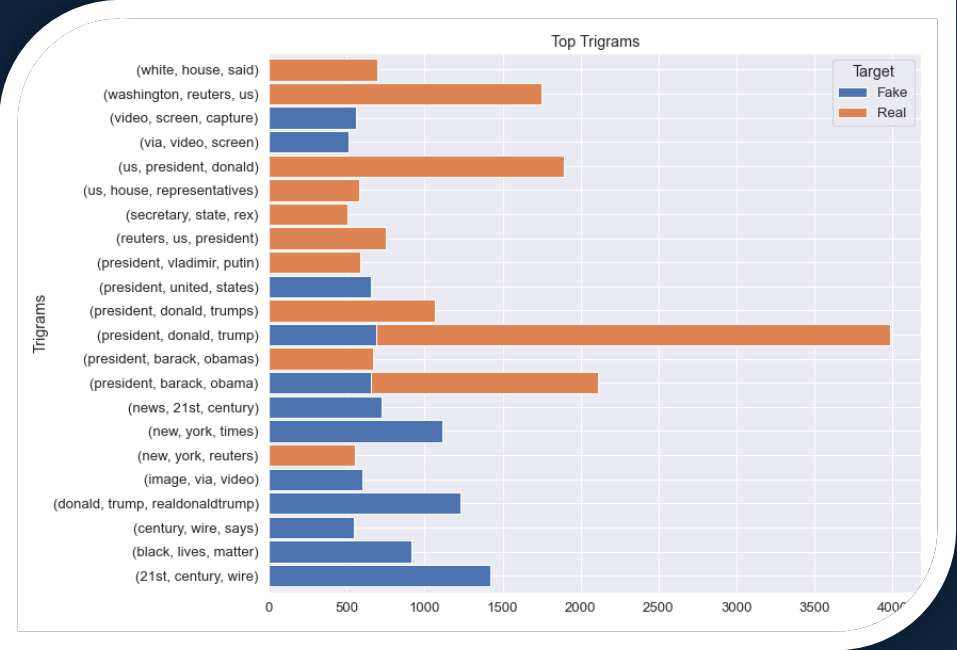
Evaluation

Type of Model	Training Accuracy
Baseline Model	87.53%
Stopwords Removed	90.76%
Stemmed	90.75%
Lemmatized	90.02%
Engineered Features	97.79%
Bigrams and Trigrams	87.93%
Bigrams and Trigrams Engineered	97.56%
Bigrams and Trigrams Stemmed Engineered	97.40%
Bigrams and Trigrams Lemmatized Engineered	97.37%
Engineered with Additional Features	97.97%

Real vs. Fake News
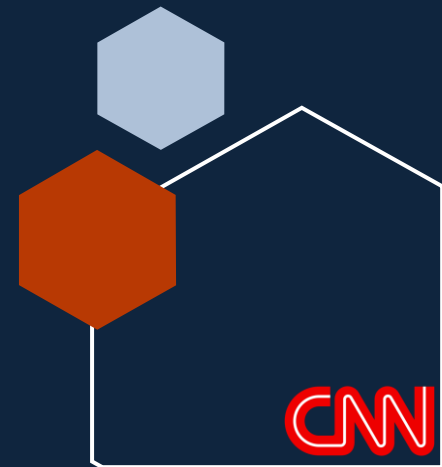
Evaluation

Type of Model	Training Accuracy	Testing Accuracy
Engineered Features	97.79%	98.05%
Bigrams and Trigrams Engineered	97.56%	97.72%
Bigrams and Trigrams Stemmed Engineered	97.40%	97.61%
Bigrams and Trigrams Lemmatized Engineered	97.37%	97.62%
Engineered with Additional Features	97.97%	98.26%

Real vs. Fake News

CNN

Recommendations & Next Steps

Use model on news articles

Analyze additional newer data

Engineer additional features

Run time analysis model

CNN

Thank You!

Kevin Gross

kevgross89@gmail.com

github.com/kevgross89