

News Media Bias: Donald Trump



Andrew Arteaga, Teresa Bui, Kyle Lee



Table of contents

01

Introduction

Objective
About the Data

02

Pre-Processing

How we cleaned the data

03

Word Analysis

Vocabulary, Sentiment
Analysis, Word Clouds

04

N-grams

Bi-grams by Counts
Negation Analysis

05

Topic Modeling

LDA
Topic Distribution by
Source



Introduction



Donald Trump remains one of the most polarizing political figures of all time. Our investigation reviews whether liberal and conservative news sources report news about Trump unbiasedly.





About our Dataset

- Columns:

1

2

3

4

5

Source	Date	Headline	Description	text
CNN	03-2024	Fact check: Trump, telling a completely fictional story...	Former President Donald Trump told an entirely fictio...	Former President Donald Trump told an entirely fictio...
CNN	03-2024	Trump warns of 'bloodbath' for auto industry and cou...	Former President Donald Trump threatened Saturday ...	Former President Donald Trump warned Saturday that...
CNN	03-2024	Will Biden and Trump debate? The brief history and u...	To debate or not to debate is a complicated question ...	A version of this story appeared in CNN's What Matter...
CNN	03-2024	Why does Biden keep mentioning January 6? Because ...	The rallies start with a recording of January 6 prisone...	The rallies start with a recording of January 6 prisone...
CNN	03-2024	Trump visits Ohio ahead of competitive Senate primar...	Donald Trump would normally have little reason to be...	Donald Trump would normally have little reason to b...
CNN	03-2024	'Everything he is saying isn't true': Congolese govern...	There is no evidence for former President Donald Tru...	There is no evidence for former President Donald Tru...
CNN	03-2024	Judge delays Trump hush money trial in New York	The judge overseeing Donald Trump's criminal hush ...	The judge overseeing Donald Trump's criminal hush ...

- 1994 rows

Text Cleaning

Corpus Cleaning Steps

- Converted all letters to lowercase
- Removed punctuation
- Removed English stop words
- Removed "cnn", "cnssrs", "cnbc"
- Stripped all whitespaces
- Stemmed document
- Removed sparse terms

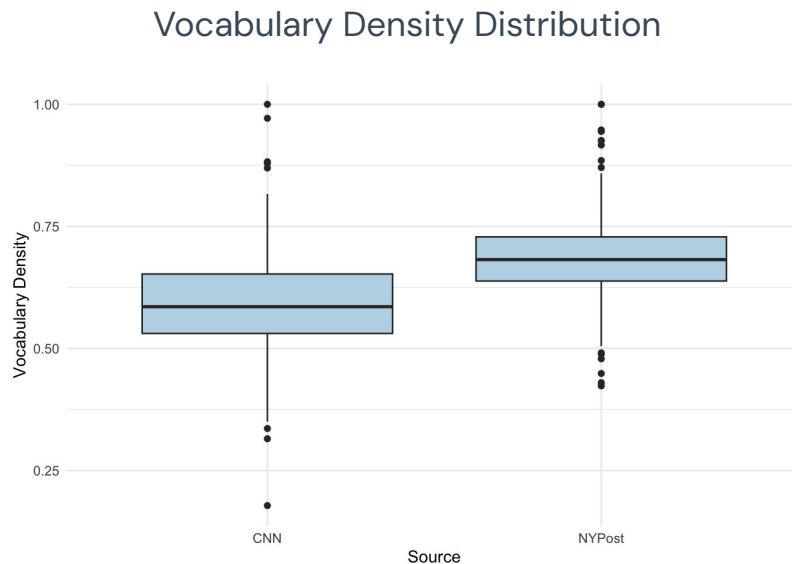
Vocabulary Break Down

CNN

- Average total terms: 527.64
- Average unique terms: 289.95
- Average vocab density: .59

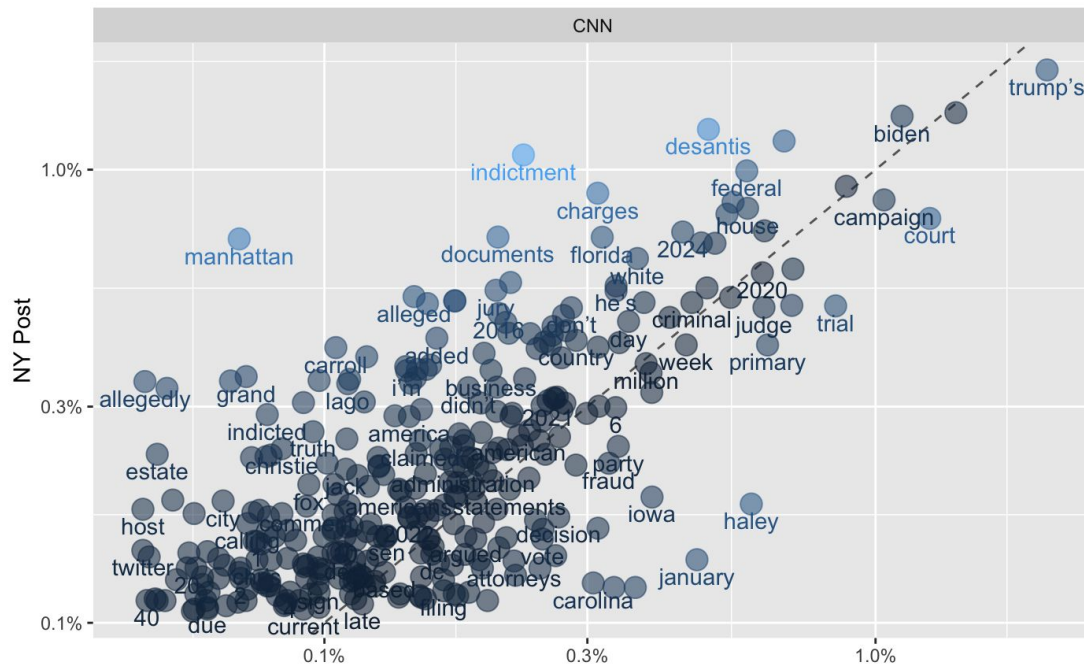
NY Post

- Average total terms: 291.30
- Average unique terms: 192.39
- Average vocab density: .6828



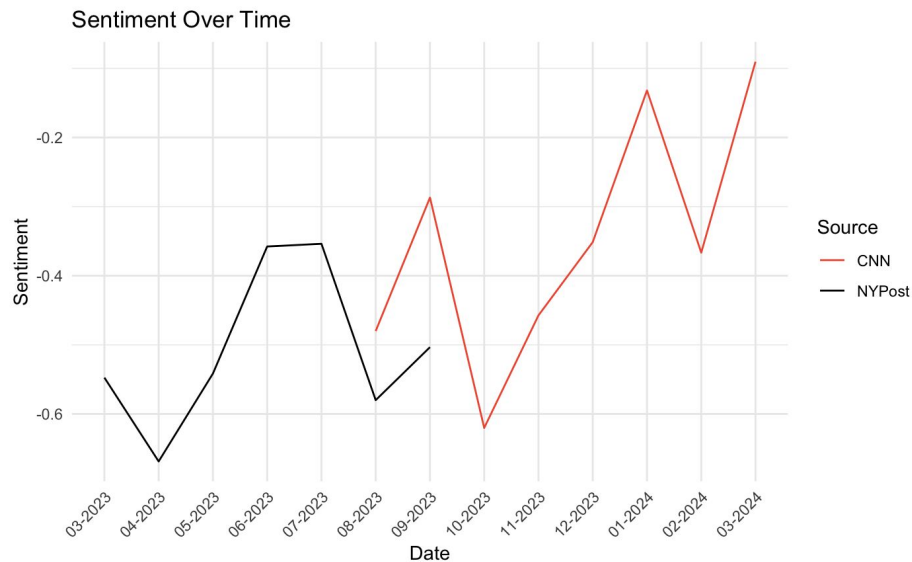
Most Frequent Word by Source





Sentiment Over Time

Generally increase over time



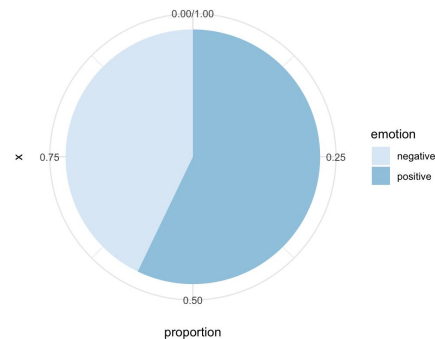
Preliminary Article Body Sentiment Overview

A quick look at the sentiment distribution for each source reveals that there is little emotional difference between each source's articles about Trump

Interestingly enough, they both lean towards a more positive emotion when writing about the former President.

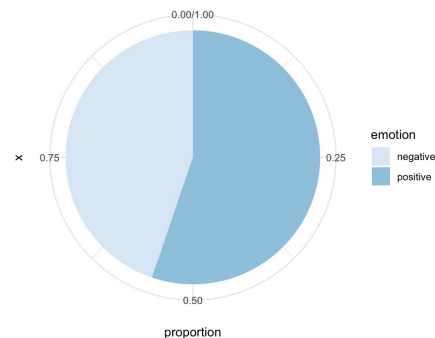
Could be because of all the court cases against him...

CNN Emotion Proportion



Positive	Negative
0.5714	0.4286

NY Post Emotion Proportion

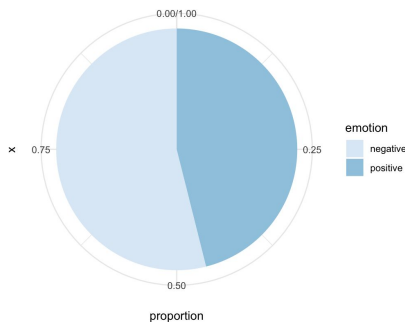


Positive	Negative
0.5522	0.4477

Headline Sentiment Overview

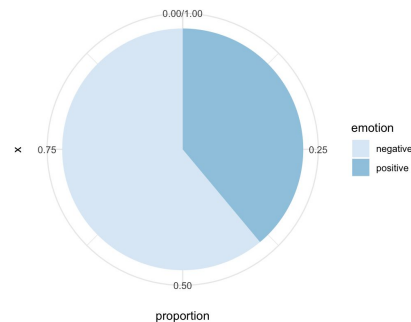
A sentiment analysis of the headline reveals that they are considerably negative especially when compared to the text to which they correspond to.

CNN Headline Emotion Proportion



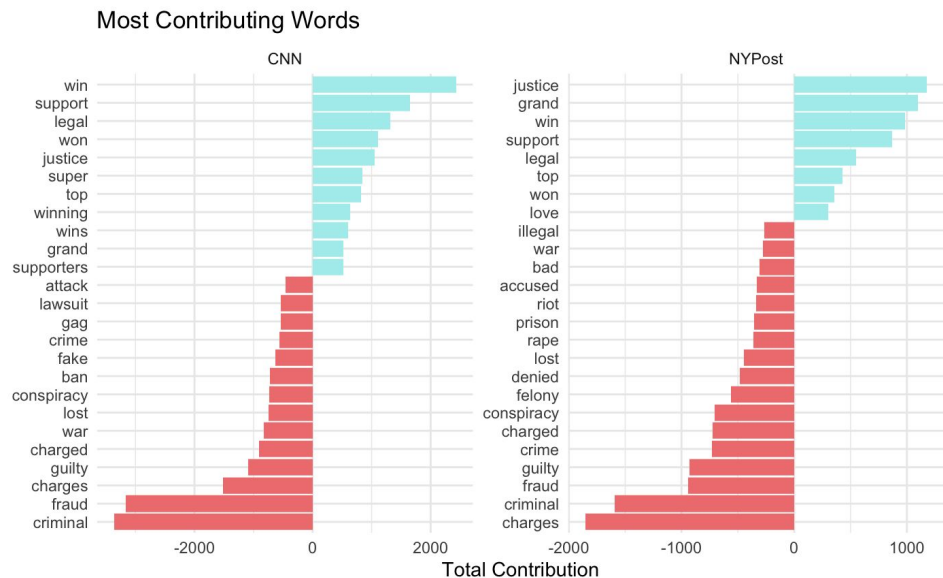
Positive	Negative
0.4608	0.5392

NY Post Headline Emotion Proportion



Positive	Negative
0.3892	0.6108

Word Contributions to Sentiment



LDA Topic Modeling

Political Campaigns

campaign dynamics, candidate messaging, voter engagement tactics, and analysis of political races

Legal Proceedings

legal actions, court cases, trials, hearings, legal strategies

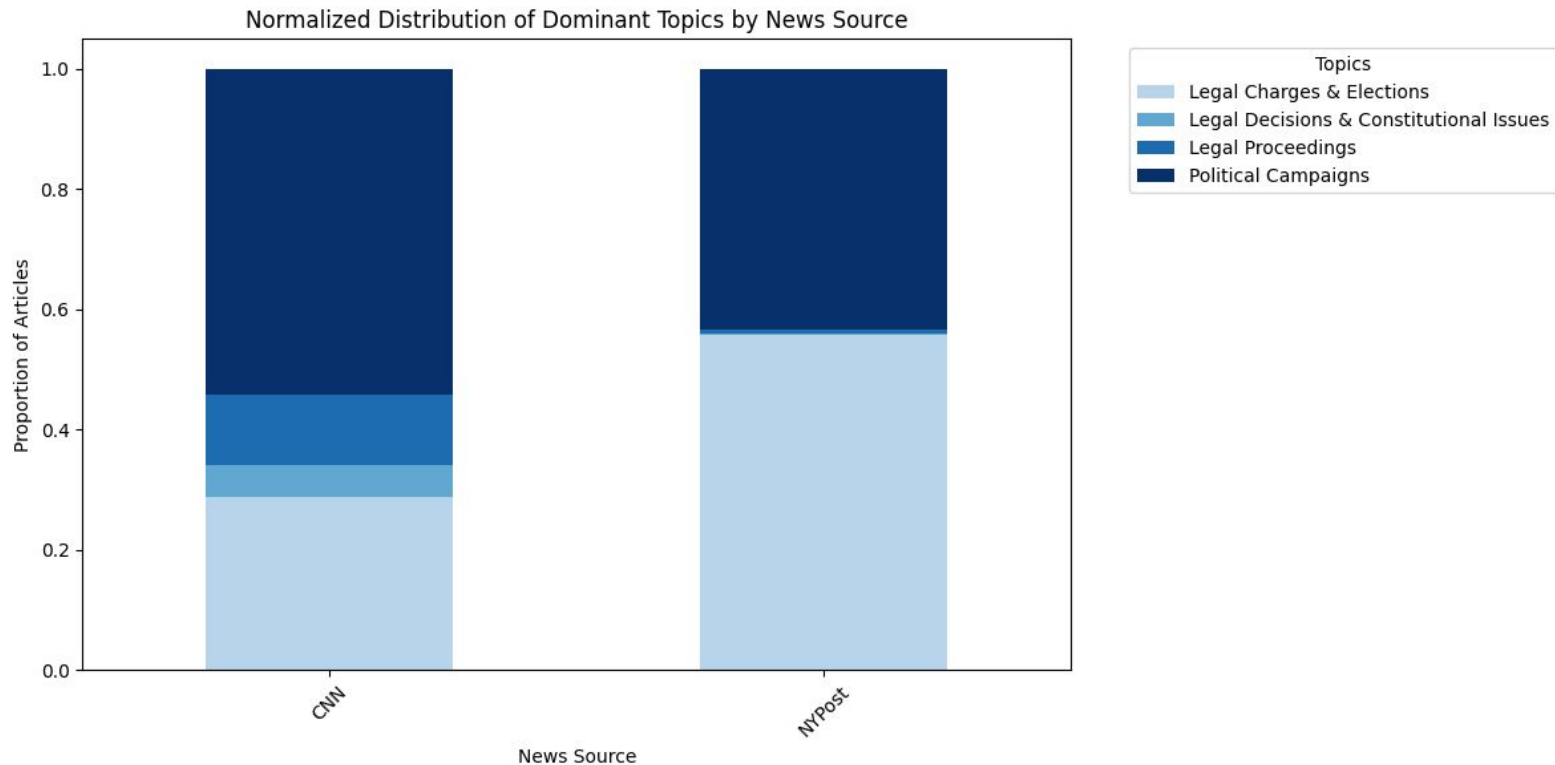
Legal Decisions & Constitutional Issues

Supreme Court rulings, constitutional amendments, and legal interpretations

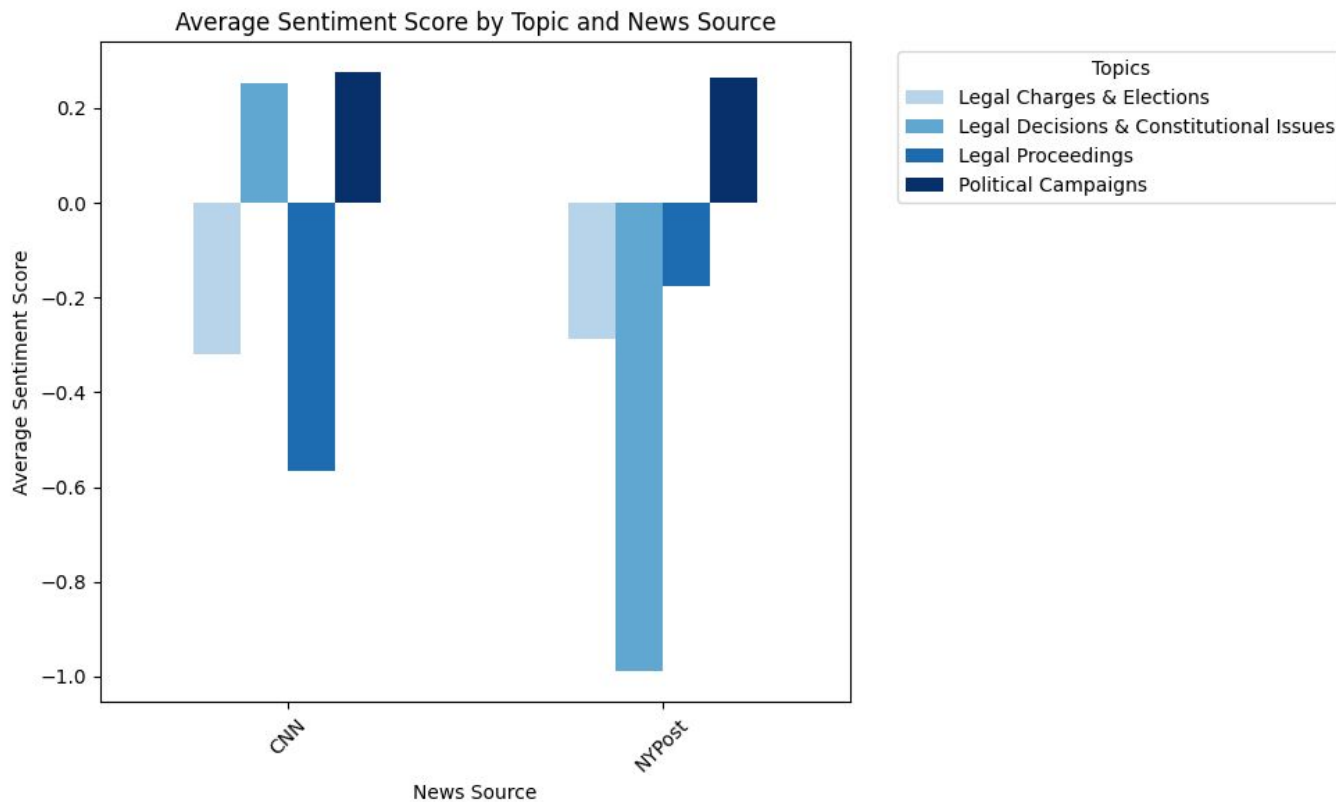
Legal Charges & Elections

legal charges against politicians or political entities, investigations impacting elections, and controversies around electoral legality

Topic Distribution by Source

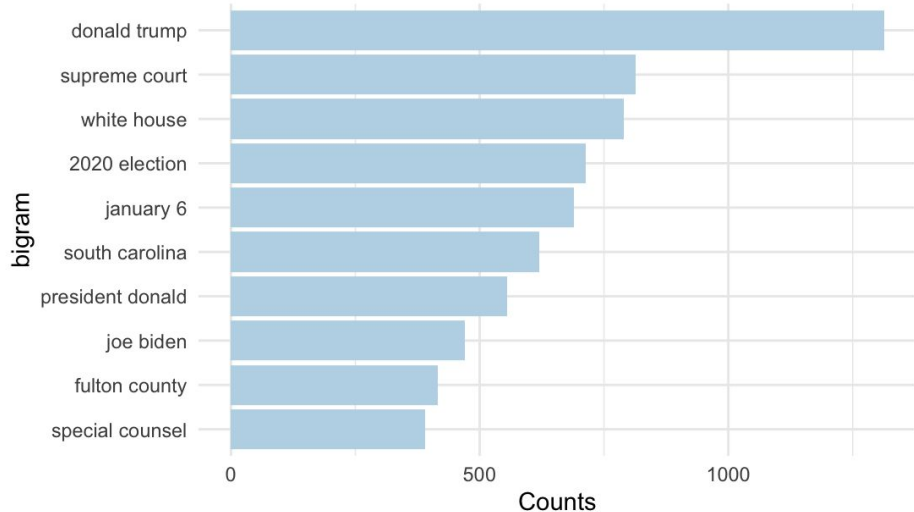


Sentiment Analysis by Topic

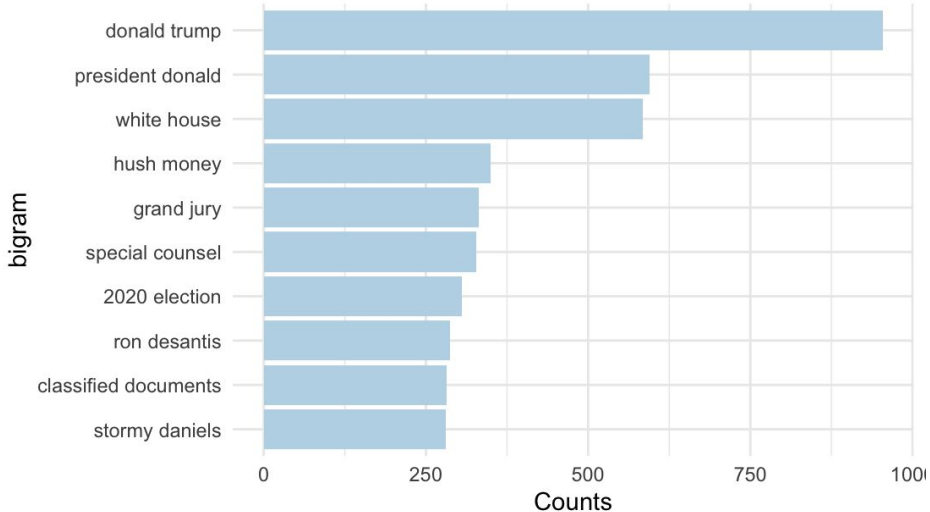


Bi-grams by Counts

CNN



NY Post

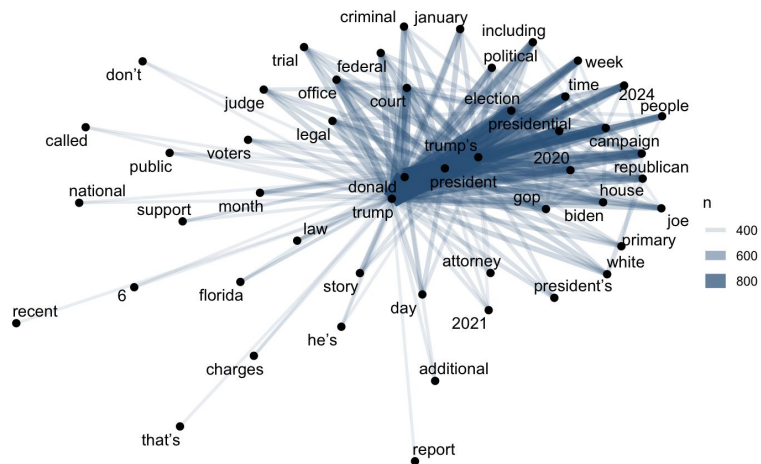


Word Correlation Plots

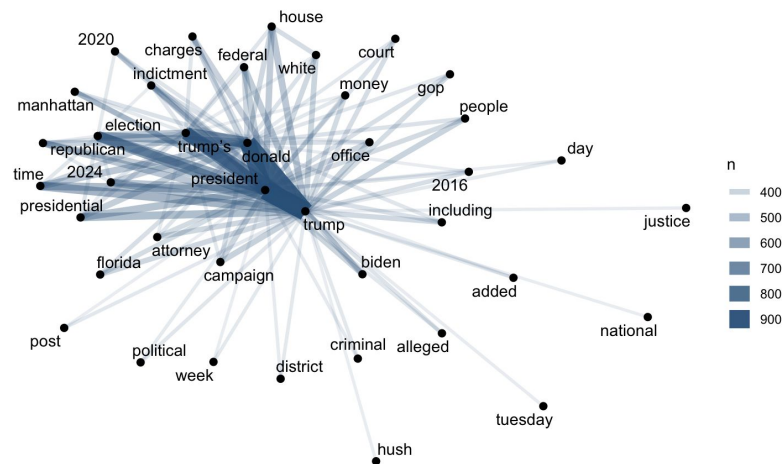
CNN

NY Post

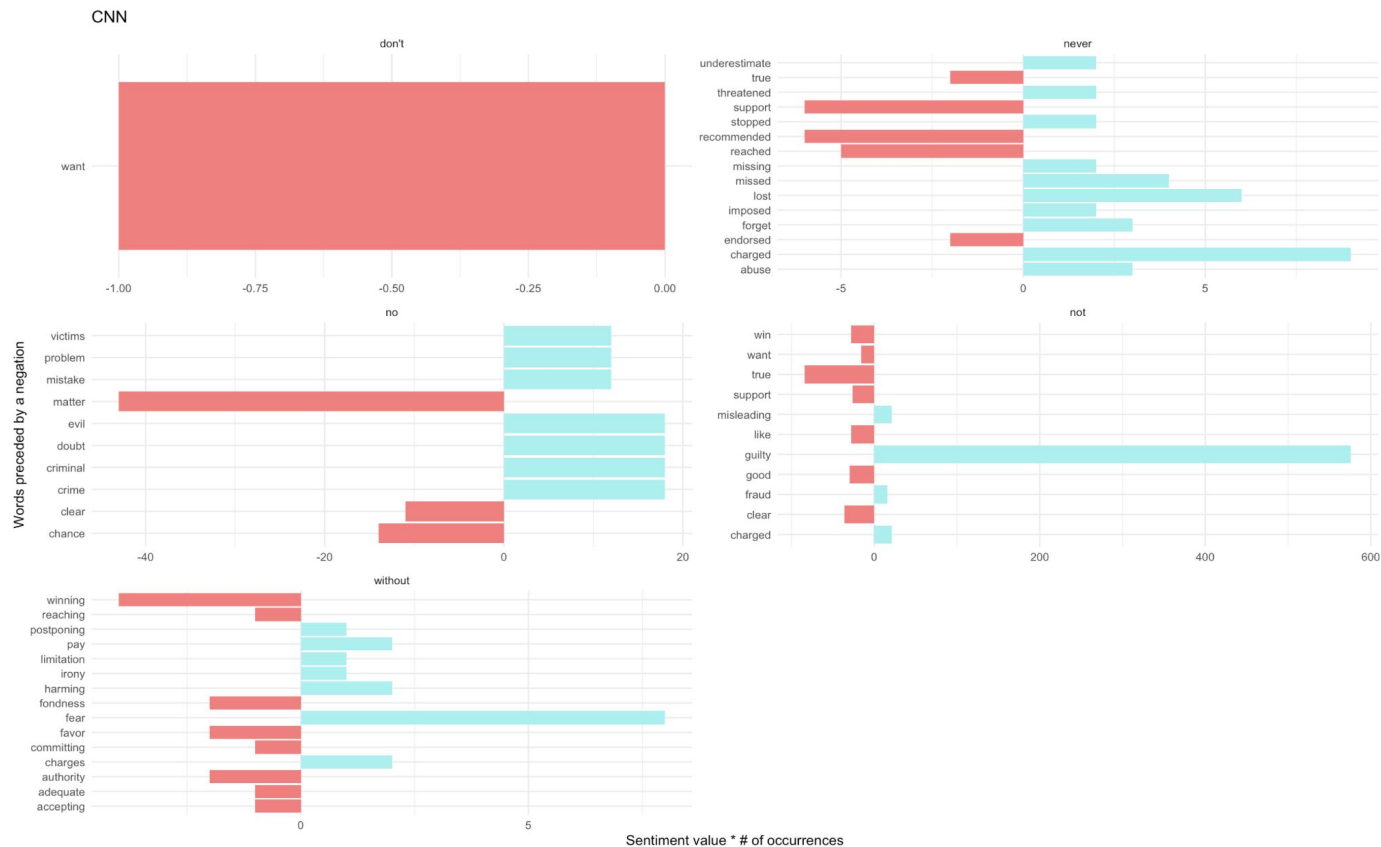
CNN



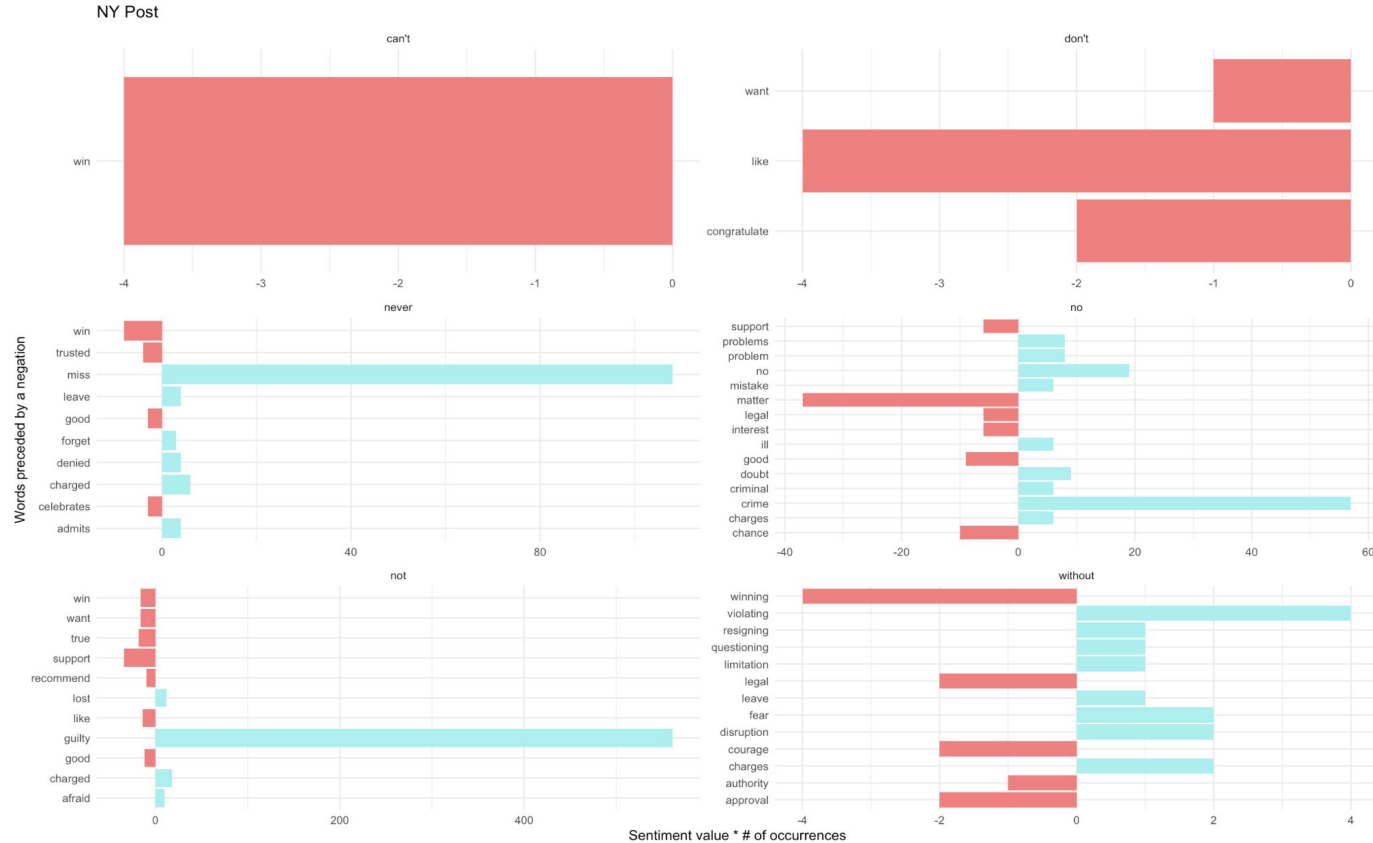
NY Post



Negation Analysis: CNN



Negation Analysis: NY Post



Classification Model / Confusion Matrix

Model

Logistic Regression

Random Forest

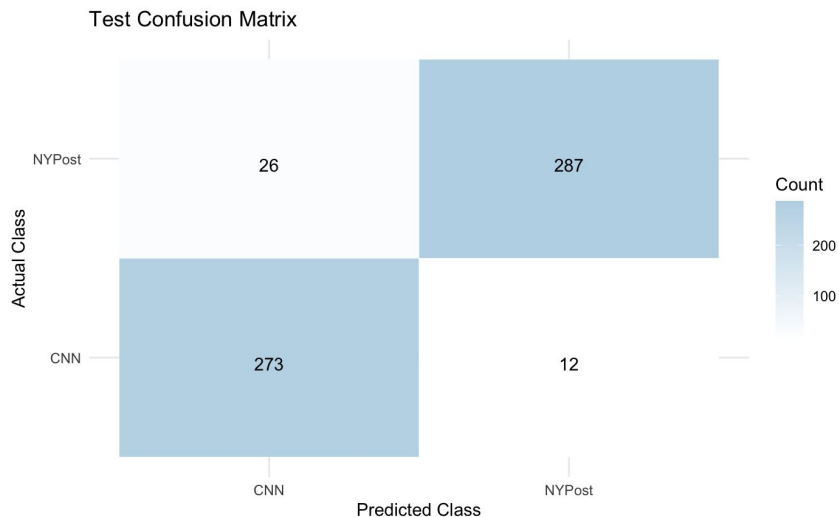
Ken's Nearest

Accuracy

92.5%

93.64%

75.75%



Thank you

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

