DO MEDIA COMPANIES DRIVE BIAS?
USING SENTIMENT ANALYSIS TO MEASURE MEDIA BIAS IN NEWSPAPER TWEETS

by
Taylor Thomsen

A capstone submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Government Analytics

Baltimore, Maryland
April, 2018

Abstract

As natural language processing tools are advancing in their use to study corpuses, it is important to harness these tools to study media content for potential media bias. Existing studies have attempted to measure political sentiment by using lexicons to determine whether political texts have a positive or negative connotation and other studies have looked at media bias by manually classifying media content. Although machine learning models are widely used to simulate human decisions about sentiment, few studies have used automated sentiment analysis on newspaper content to measure media bias. This paper uses an out-of-the-box sentiment analysis model on several newspapers' tweets from four major media companies during the month leading up to the 2016 presidential election. The sentiment analysis results in a sentiment score for each tweet mentioning Republican or Democratic keywords. Overall, this paper finds that only some newspapers had significant differences in the sentiment scores for Republican and Democratic tweets. Additionally, Republican and Democratic sentiment scores were not significantly different between media companies, showing that the companies that own the newspapers may not be driving biased content of the individual newspapers.

Table of Contents

Tables and Figures

1 Introduction

A 2017 Gallup study states that 62% of Americans believe that the news media favors one political party over the other.[1] During the 2016 presidential election, Donald Trump was outspoken about his belief that the media favored his opponent, Hillary Clinton. He and many consumers may subjectively call a news source biased or untrustworthy, but what can we do to objectively evaluate a news source? It is important to have quantitative measures of media bias in order to add hard evidence to the ongoing conversation.

How do we know what media sources to believe and which news sources to cautiously consume? In a world that seems increasingly divided by partisan politics, it is important to be well informed and to not blindly let the media influence one's opinions. The news sources consumers go to play a role in political polarization and consuming biased news can further this divide between parties and between citizens.[2] It is important for the media to provide well-rounded news and not affirm the beliefs of its journalists or its readers. This relationship is two-sided. Media outlets must be conscious of providing unbiased media and consumers must be skeptics of news reporting. But how do consumers know who to trust?

Quantitative measures of media bias can help people be more responsible media consumers. It helps them to know which sources to trust and which sources to be skeptical of. Many researchers have attempted to create tools to help measure media bias. There is not one specific way that media bias should be measured, so these tools vary widely. Some researchers have counted what organizations are cited in newspapers and compared these citations to what organizations members of Congress usually cite in their speeches and

---

[1] *Gallup*, *"Six in 10 in U.S. See Partisan Bias in News Media,"* http://news.gallup.com/poll/207794/six-partisan-bias-news-media.aspx (accessed April 13, 2018).
[2] *Pew Research Center*, *"Political Polarization and Media Habits,"* http://www.journalism.org/2014/10/21/political-polarization-media-habits/ (accessed April 14, 2018).

writing.[3] Other research measures the number of positive and negative stories published on a specific topic across newspapers.[4] Using many different tools to measure media bias can result in different results regarding which media outlets are biased and in what direction. By combining these tools and results, one can begin to gather how some media outlets are biased in the places that research results overlap.

Sentiment analysis describes a machine learning classification problem that categorizes a given text based on how positive or negative the contents of the text are. Sentiment analysis has been employed to study public opinions on products and companies as well as to study political texts created by members of Congress and public opinions of politicians. Although there are several studies that use sentiment analysis to study different topics, sentiment analysis has yet to be utilized to study media bias.

This paper aims to understand whether or not media sources produce biased content by applying natural language processing and sentiment analysis as a method of measuring media bias. The tools for natural language processing and sentiment analysis have become more sophisticated over the years and have found more applications, but there is still room for research on applying sentiment analysis to studying media bias.

This paper combines this idea with the concept of using Twitter as a text corpus to study Twitter content produced by newspapers. To do so, keyword-based sentiment analysis was performed on the tweets of the five largest newspapers from four media companies (20 newspaper accounts total) leading up to the 2016 presidential election. This means that tweets containing one or more partisan keywords were identified and put through the sentiment analysis model. The sentiment analysis results showed that there was no significant

---

[3] Groseclose, Tim and Jeffrey Milyo, "A Measure of Media Bias," *The Quarterly Journal of Economics* 120, no. 4 (2005).

[4] Soroka, Stuart, "The Gatekeeping Function: Distributions of Information in Media and the Real World," *Journal of Politics* 74, no. 2 (2012): 515.

difference in the political sentiment of tweets from different media companies and only a few newspapers showed different sentiment in Republican and Democratic tweets. Although this study finds little evidence that the newspapers of interest lean towards one side of the political spectrum, more research should be conducted to use the powerful abilities of sentiment analysis in detecting media bias. The next section includes a more in depth review of existing literature, followed by a section describing the details of the sentiment analysis performed on newspaper tweets, and finally a conclusion that will propose areas of future research.

2 Literature Review

2.1 Measuring Media Bias

Most studies measure bias in relation to other news outlets or in relation to politicians by first giving politicians some type of ideological score and then comparing content created by those politicians to news content. Media bias can be measured in several ways, but is commonly divided into three types of media bias. The literature on media bias has different terminology for types of bias, but they can be summarized in these three terms: coverage bias, agenda bias, and tonality bias.[5] Coverage bias refers to how equally both parties or both sides of a given issue are covered.[6] Agenda bias is also referred to as "gatekeeping" or "selectivity bias" and occurs when the media chooses to print only certain stories. This also includes when media outlets give more attention to the preferred issues of one party over another party.[7] Tonality bias occurs when members of the media, including writers, editors, or owners, insert their own opinions into the content they produce. It

[5] D'Alessio, Dave and Mike Allen, "Media Bias in Presidential Elections: A Meta-Analysis," *Journal of Communication* 50, no. 4 (2000): 135.
[6] Eberl, Jakob-Moritz, Hajo G. Boomgaarden, and Markus Wagner, "One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences," *Communication Research* 44, no. 8 (2017): 1128.
[7] Ibid., 1128-1129.

includes how politicians or issues are evaluated by news sources and the sentiment in which politicians or issues are described.

One widely cited study of media bias is that of Groseclose and Milyo from 2005 in which these researchers measured the amount of times think tanks were cited in the media and then compared these statistics to the amount of times members of Congress cited the same think tanks.[8] The think tank citations were translated into an ideological score by comparing a media outlet with the Americans for Democratic Action (ADA) annual Liberal Quotient of Congress members. The ADA Liberal Quotient is a score from 0 to 100 where 100 means the member of Congress completely agrees with ADA policies and 0 means the member of Congress completely disagrees with ADA policies.

The results of this study showed that most media outlets had a liberal bias, citing think tanks that were more commonly cited by liberal members of Congress. Exceptions to this result were *Fox News Special Report* and the *Washington Times*. The more neutral news outlets consisted of *PBS NewsHour*, *CNN Newsnight*, and *Good Morning America*. *CBS Evening News* and the *New York Times* received the most liberal scores.[9] This comparison research method is effective because it does not require humans to make subjective decisions on media content and instead relies on an objectively measurable approach.

Another popular way to measure media bias is to have human coded evaluations of sentiment in the news. For example, in 2004, Lott and Hassett scored economic headlines from 389 American newspapers between 1991 and 2004 on tone.[10] The tone of the headlines was compared between periods of time when a Democrat was president versus when a Republican was president. The study found that newspapers were more likely to have more

[8] Groseclose and Milyo, "Media Bias," 1191–1237.
[9]Ibid., 1204–1207.
[10] Lott, John R. and Kevin A. Hassett, "Is Newspaper Coverage of Economic Events Politically Biased?" *Public Choice* 160, no. 1-2 (2004): 65–108.

positive coverage of the same economic news when a Democrat was president than when a Republican was president.[11]

A study conducted by Genztkow et al. in 2006 used a predefined list of words for sentiment coding. The researchers studied how newspapers covered two bribery scandals: the Credit Mobilier scandal of the 1870s — in which the Credit Mobilier of America construction company overcharged railroad rates and bribed politicians — and the Teapot Dome scandal of the 1920s — in which Navy petroleum reserves were leased to private oil companies at lower rates. The researchers compared the coverage of the two scandals by counting negative words such as "slander" and "liar" against positive words such as "honest" and "honorable." The study found that coverage of the Teapot Dome scandal was less biased because of the changing expectations from the media at the time.[12]

2.2 Measuring Bias with Computer Automation

Manually coding sentiment does not scale well and can add limitations to the reach of a study.[13] For this reason, some sentiment research focuses on creating reusable lexicons for coding sentiment, such as Soroka and Young's Lexicoder Sentiment Dictionary (LSD), which combines three existing sentiment lexicons to create one large dictionary of positive and negative words to be used for sentiment analysis.[14]

Lexicon dictionaries are used in both supervised and unsupervised machine learning models to automate sentiment analysis. Supervised machine learning models take human coded information (sometimes in the form of a lexicon) and predict how a human would

---

[11] Lott and Hassett, "Newspaper Coverage of Economic Events."
[12] Gentzkow, Matthew, Edward L. Glaeser and Claudia Goldin, "The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered," *National Bureau of Economic Research Corruption and Reform: Lessons from America's Economic History* (2006).
[13] Haselmayer, Martin, and Marcelo Jenny, "Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding," *Quality & Quantity* 51, no. 6 (2017): 2625-2626.
[14] Soroka, Stuart and Lori Young, "Affective News: The Automated Coding of Sentiment in Political Texts," *Political Communication* 29 (2012), 205–231.

code further examples based on the examples that have been already human coded. The purpose of automated models is to scale research by giving the ability to code more data points or to classify more sentiment examples than previously possible with human coding alone. Today many sentiment machine learning models exist as APIs that are available for research use.

2.3 Sentiment Analysis on Twitter Data

Twitter is a microblogging service and social media platform where users can compose short tweets. In the last quarter of 2016, Twitter had 318 million active monthly users worldwide[15], of which 67 million were in the United States.[16] Twitter data can be obtained through the Twitter APIs, which provide access to archived data as well as real-time data. Twitter has become a popular tool for natural language processing due to data availability and ease of access. It is popularly used in natural language processing research to study sentiment towards topics that are trending on Twitter or to study support for political issues and candidates.

Twitter makes it easy to gather samples of text from many different authors at many points in time, or even in real time, and sort these text snippets by topic or user. The main differences between using tweets as a corpus over a larger text corpus are the length of each text corpus, the availability of tweets, and grammatical structure.[17]

[15] "Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 4th Quarter 2017 (In Millions)," *Statista*, last modified February 2018, https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.
[16] "Number of Monthly Active Twitter Users in the United States From 1st Quarter 2010 to 4th Quarter 2017 (In Millions)," *Statista*, last modified February 2018, https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/.
[17] Chong, Wie Yen, Bhawani Selvaretnam and Lay-Ki Soon, "Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets," (Paper presented at the *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*, 2014).

First, tweets are much shorter than other common corpuses like political texts or newspaper articles. Tweets have been historically limited to 140 characters and in 2017 the character limit was expanded to 280 characters. This is different from other common corpuses such as the political texts from Congress members used in Groseclose and Milyo's study.[18]

Secondly, tweets are available in large quantities and are therefore useful for training machine learning models. Go et al. used Twitter as a source for testing the accuracy of several machine learning models on sentiment analysis. The study tested the sentiment of Tweets towards certain products and brands with the goal of allowing users to research brands before they make a purchase and for companies to keep track of attitudes toward their products.[19]

Lastly, tweets have different sentence structure than traditional text corpuses due to their social nature and length restrictions. Tweets often have more informal language such as acronyms or slang and can include hyperlinks, pictures, hashtags, and emoticons. These informal properties add noise to some machine learning problems and in other cases they can be used as a classification feature and can contribute to training the model. For example, in the aforementioned Go et al. study, emoticon data reduced the accuracy of two machine learning models and had no affect on one model when used as a feature.[20] In this paper, informal tweets are less likely to be a challenge because tweets are solely taken from newspaper accounts, which are generally more professional than tweets from the average Twitter user.

---

[18] Groseclose and Milyo, "Media Bias."
[19] Go, Alec, Richa Bhayani, and Lei Huang, "Twitter Sentiment Classification Using Distant Supervision," *CS224N Project Report, Stanford* 1, no. 12 (2009): 1.
[20] Go, Bhayani, and Huang, "Twitter Sentiment Classification," 2.

Pak and Paroubek's study on evaluating Twitter as a corpus for sentiment analysis shows that Twitter is a helpful source for sentiment analysis because the corpus is large, constantly expanding, and provides a variety of data. With the number of tweets growing daily, you can get tweets from a variety of authors – celebrities, scientists, politicians, CEOs, average Joes, etc. – on various topics and in different languages.[21] Twitter users mainly tweet to express their opinions about a topic, leading researchers to believe that Twitter can be a rich corpus for sentiment analysis. Given these advantages, Pak and Paroubek created an automated way to collect Twitter data, performed statistical linguistic analysis on the data, built a multinomial Naïve Bayes sentiment classifier, and tested the accuracy of the sentiment classifier.

In the evaluation of the sentiment classifier, the researchers discovered that part of speech (POS) tagging was among the best features for the sentiment classifier because it helped distinguish between objective and subjective tweets. Common or proper nouns appeared more often in objective tweets while subjective tweets included more personal pronouns and utterances.[22] According to part of speech tagging guidelines, utterances include word such as: "'my' (as in 'My, what a gorgeous day'), 'oh,' 'please,' 'see' (as in 'See, it's like this'), 'uh,' 'well,' and 'yes,' among others."[23] The importance of POS tagging to the sentiment classifier shows the uniqueness of using tweets as a corpus over a more traditional corpus.

Political scientists and machine learning engineers have also evaluated Twitter as a source for studying political sentiment. Waykar, Wadhwani, and More took the work Go et

---

[21] Pak, Alexander, and Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," (Paper presented at the International Conference on Language Resources and Evaluation, Valletta, Malta, May 17-23 2010).
[22] Pak and Paroubek, "Twitter as a Corpus," 1322.
[23] Santorini, Beatrice, "Part-of-Speech Tagging Guidelines for the Penn Treebank Project," *Linguistic Data Consortium* (1990): 3.

al. completed on using Twitter as a source of sentiment on products or companies and applied it to determining the political leanings of Twitter users.[24] To do this, sentiment was measured in relation to a political party or candidate as the query term instead of a product or company. The findings were similar to that of Go et al., but within a political context instead of a marketing context.

Furthering the research of political sentiment analysis on Twitter data, O'Conner et al. suggests that political sentiment analysis on tweets could even be a faster and inexpensive replacement for traditional polling.[25] This study used query term-based sentiment analysis to replicate consumer confidence polls (keywords: "economy," "job," and "jobs"), presidential approval (keyword: "Obama"), and elections (keywords: "Obama" and "McCain"). The sentiment results on the word "jobs" showed a similar pattern as the Gallup Daily and Michigan ICS consumer confidence polls. One interesting finding was that the tweet sentiment results were quicker to show changes in consumer confidence sentiment than the polls were and the Twitter results somewhat acted as a predictor of what direction the polls were heading. The consumer confidence polls during the time frame studied had a rather simple pattern — an upward trend until 2008, a downward trend during 2008, and a rebound in spring of 2009 — which may explain why other query-based sentiment analyses did not correlate as closely with polls. Overall, this study showed that there may be an important link between Twitter sentiment and polls, but this relationship is complicated.

With this knowledge, it is worth using political sentiment from tweets to learn about political sentiment in the media.

---

[24] Waykar, Pranav, Kailash Wadhwani, and Pooja More, "Sentiment Analysis in Twitter Using Natural Language Processing (NLP) and Classification Algorithm," *International Journal of Advanced Research in Computer Engineering & Technology* 5, no. 1 (2016): 79-81.
[25] Brendan O'Connor et al., "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." *International Conference on Web and Social Media* 11 (2010): 122.

3 Data and Methods

3.1 Descriptive Statistics

For the following analysis, data was obtained from the Twitter API using a Python library called Tweepy. Tweets were gathered from newspapers belonging to four large media companies: Digital Media First, Gannett, McClatchy Co., and Tronc Inc. Tweets were gathered from five Twitter accounts for each media company during the month leading up to the 2016 presidential election. Before filtering, these 20 newspaper Twitter accounts produced roughly 29,000 tweets leading up to the election. The tweets were narrowed down by content and only political tweets were kept for analysis, resulting in 4,215 political tweets. In order for a tweet to be considered political, it needed to contain one of the following words: "Democratic", "Hillary", "Clinton", "Democrat", "Clinton's", "Tim," "Kaine," "Kaine's," "Republican", "Donald", "Trump", "GOP", "Trump's," "Mike," "Pence," or "Pence's." Note that the keyword "Tim" was included with a space following to make sure words like "time" were not included. The analysis includes tweets even when a keyword is only found in a hashtag of the tweet.

Political tweets made up 14.5% of the total tweets produced by the observed newspapers. The tweets were flagged into three categories based on which words the tweet contained. The three categories were Republican Party related tweets, Democratic Party related tweets, or tweets mentioning both the Democratic and Republican Parties. Table 1 shows the total number of tweets for each category. These totals show that there were almost twice as many Republican tweets than Democratic tweets leading up to the election.

Table 1: Number of Tweets Classified as Republican, Democratic, or Both

Category	Number of Tweets
Republican Tweets	3041
Democratic Tweets	1654
Both Republican and Democratic Tweets	701

Table 2 shows the number of tweets each newspaper produced during the timeframe for each of the three categories. Notably, *USA Today*, the *New York Daily News*, and the *LA Times* were the most active on Twitter, tweeting 1,023, 820, and 768 tweets during that month, respectively.

Table 2: Tweets Classified as Republican, Democratic, or Both by Newspaper

Company	Newspaper Account	Republican Tweets	Democratic Tweets	Both Republican and Democratic Tweets
Digital First Media		397	189	93
	@denverpost	126	62	26
	@ladailynews	91	36	25
	@mercnews	150	69	30
	@nhregister	13	10	4
	@ocregister	17	12	8
Gannett		929	485	274
	@Enquirer	85	39	14
	@USATODAY	597	311	127
	@azcentral	93	74	23
	@indystar	77	24	19
	@journalsentinel	77	37	24
McClatchy Co.		385	274	96
	@KCStar	135	76	31
	@MiamiHerald	145	128	36
	@startelegram	46	27	11
	@theobserver	57	42	18
	@thestate	2	1	0
Tronc Inc.		1326	703	304

@NYDailyNews	509	230	91
@SunSentinel	82	64	24
@chicagotribune	289	153	76
@latimes	433	242	106
@sdut	13	14	7

3.2 Natural Language Toolkit and VADER Sentiment Model

The Python Natural Language Toolkit was used for preprocessing and examining Twitter data and the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment model was used for all sentiment analyses. The VADER sentiment model is a sentiment classifier that results in a sentiment confidence level.[26] To calculate sentiment, VADER uses both lexicon analysis and rule-based sentiment analysis to provide a measure of how negative, positive, and neutral a body of text is. The VADER measure used in this paper is the aggregate sentiment measure, which provides one sentiment score in which -1 is the most negative, 0 is neutral, and 1 is the most positive.

4 Key Findings

Using the VADER sentiment model, sentiment analysis was performed on the remaining political tweets. Table 3 shows the average sentiment statistic for each newspaper in the three previously mentioned categories. The average sentiment scores show that for most papers, tweets were fairly close to neutral. The *Chicago Tribune* of Tronc Inc. had the most negative scores at -0.12, -0.12, and -0.06 for Republican related tweets, Democratic related tweets, and both party related tweets respectively. The *Orange County Register* of Digital First Media had the most positive sentiment scores with 0.07 for Republican Party tweets

[26] Hutto, C.J. and E.E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Eighth International Conference on Weblogs and Social Media* (2014).

and 0.14 for Democratic Party tweets. The newspaper with the largest difference between Republican Party and Democratic Party tweets was the *Indy Star* of Gannett with 0.02 for Republican Party tweets, -0.11 for Democratic Party tweets, and 0.08 for tweets mentioning both the Republican and Democratic Parties.

Table 3: Average Sentiment Scores

Company	Newspaper Account	Republican Tweets	Democratic Tweets	Both Republican and Democratic Tweets
Digital First Media		-0.004	0.002	0
	@denverpost	-0.02	-0.06	0.02
	@ladailynews	0.01	0.01	0.03
	@mercnews	-0.02	-0.07	-0.04
	@nhregister	-0.06	-0.15	0.02
	@ocregister	0.07	0.14	-0.03
Gannett		-0.03	-0.01	-0.09
	@Enquirer	-0.03	-0.12	-0.16
	@USATODAY	-0.06	0.05	-0.04
	@azcentral	-0.06	0.03	-0.06
	@indystar	0.02	-0.11	0.08
	@journalsentinel	-0.04	0.10	0.09
McClatchy Co.		-0.12	-0.03	-0.08
	@KCStar	-0.01	-0.03	-0.08
	@MiamiHerald	-0.02	0.004	-0.03
	@startelegram	-0.07	0.01	-0.17
	@theobserver	-0.09	-0.002	-0.10
	@thestate	-0.42	-0.15	-
Tronc Inc.		-0.10	-0.02	-0.01
	@NYDailyNews	-0.11	-0.03	0.02
	@SunSentinel	-0.11	0.02	-0.09
	@chicagotribune	-0.12	-0.12	-0.06
	@latimes	-0.05	0.11	-0.03
	@sdut	-0.10	-0.10	0.04

4.1 Comparing Media Company Sentiment Scores

A one-way ANOVA test was run to test the differences of means between the four media company sentiment scores. The test resulted in an F-value of 2.50 and a p-value of 0.06, allowing one to conclude that there is not a statistically significant difference between the sentiment scores across media companies at a 95% confidence level. Figure 1 shows the sentiment spread of tweets for each observed company.

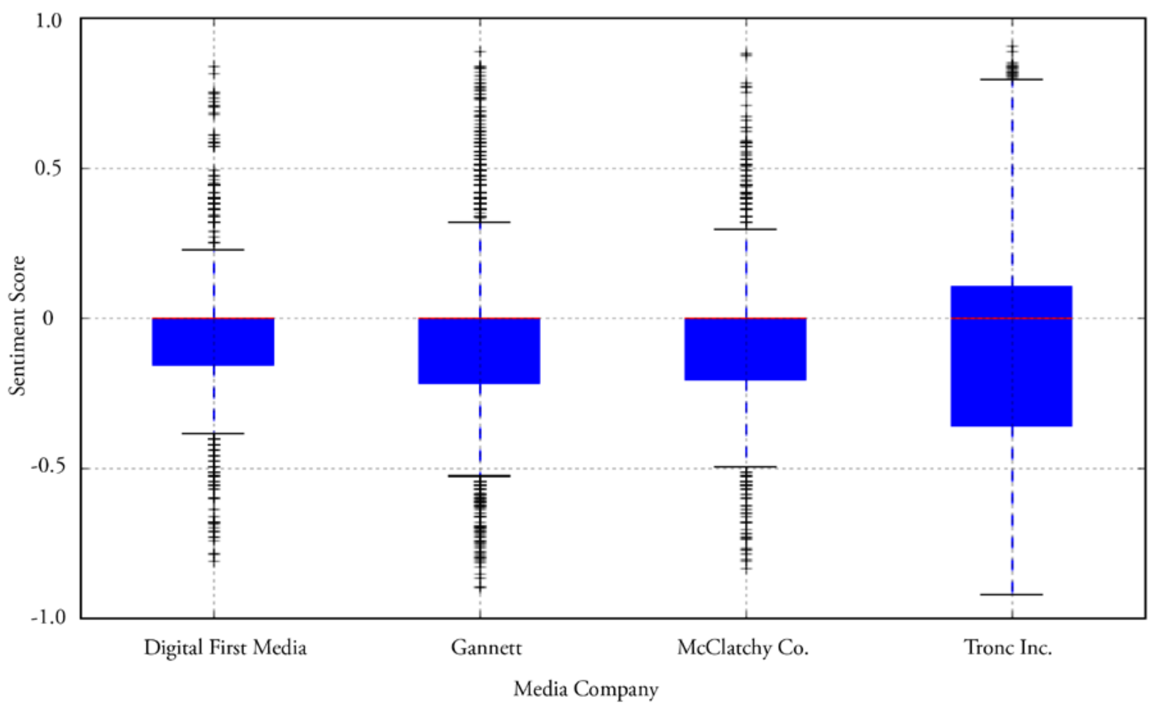

Figure 1: Sentiment Scores Across Media Companies

4.2 Comparing Newspaper Sentiment Scores

The *Orange Country Register* of Digital First Media had the highest average sentiment score with a 0.05 and the *Chicago Tribune* of Tronc Inc. had the lowest average sentiment score of -0.11. An ANOVA test between these two newspapers resulted in an F-value of 3.26 and a p-value of 0.07. The results of the ANOVA show that there is not a statistically significant difference between these two newspapers' average tweet sentiment scores at a

95% confidence level. The ANOVA test comparing these two newspapers' Republican tweet sentiment scores resulted in an F-value of 1.78 and a p-value of 0.18, while the ANOVA for the newspapers' Democratic tweets resulted in an F-value of 1.03 and a p-value of 0.31. Looking at the difference in tweet sentiment regarding one party also did not show a statistically significant difference.

From examining the average sentiment scores within categories, two Digital First Media newspapers show a notable difference in sentiment scores for tweets mentioning the Democratic Party. The *New Hampshire Register* had a sentiment score of -0.15 and the *Orange County Register* had a score of 0.14. An ANOVA test of these two papers' total tweets produced an F-value of 1.61 and a p-value of 0.21. This test is also not statistically significant and therefore the null hypothesis is retained. An ANOVA test of the Democratic tweets only produced by these two newspapers resulted in an F-value of 0.98 and a p-value of 0.33. An ANOVA test of the newspapers' Republican tweets only resulted in an F-value of 0.28 and a p-value of 0.60.

From the statistical tests performed, there is no reason to believe that media companies drive the content of their newspapers in one direction or another because there was no significant difference between average sentiment scores across companies. Furthermore, the largest variance between sentiment scores for tweets related to the Democratic Party was observed from newspapers operating under the same overarching company.

4.3 Republican Versus Democratic Tweet Sentiment Scores Overall

Without grouping tweets into newspapers or media companies, a difference of means ANOVA was conducted on the sentiment scores of all Republican and Democratic tweets. The average compound sentiment score for all Republican tweets was -0.06 and the average sentiment score for all Democratic tweets was -0.01. The ANOVA test produced an F-value of 0.12 and a p-value of 0.73, showing that although the average Republican sentiment score was slightly more negative than the Democratic sentiment score, this was not a statistically significant difference.

4.4 Republican Versus Democratic Tweet Sentiment Scores at the Newspaper Level

Figure 2 shows the distribution of sentiment scores for Democratic and Republican tweets for each newspaper. From the box plot, it is apparent that most newspapers lean towards slightly negative tweets, but there are also several outliers outside of the minimum and maximum markers.

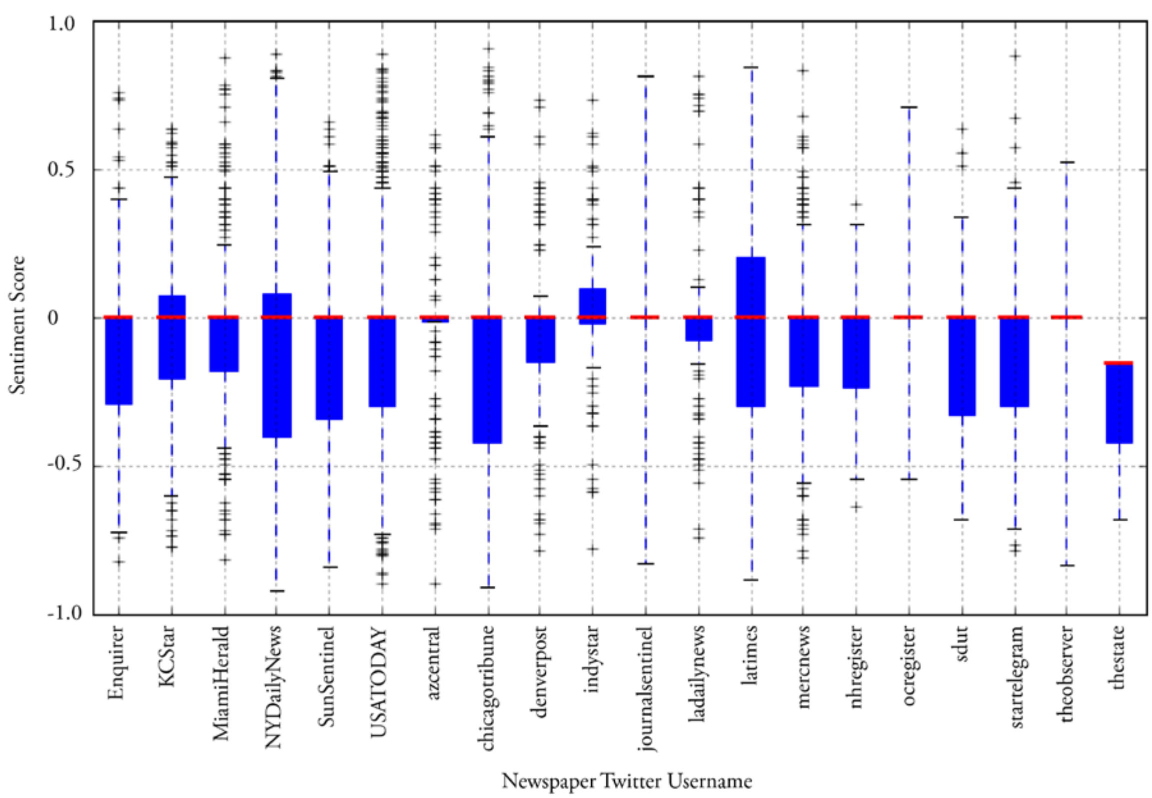

Figure 2. Sentiment Scores Across Newspapers

ANOVA tests were run at the newspaper level to test whether there was a significant difference between the average Democratic tweet sentiment score and average Republican sentiment score for each newspaper. Table 4 shows the results from these ANOVA tests. The *LA Times*, *USA Today*, and the *New York Daily News* were the only newspapers that had a significant difference between Republican and Democratic sentiment scores. Each of these three newspapers sentiment scores show a slightly more positive tone towards Democrats, although the *New York Daily News* still had a negative Democrat sentiment score of -0.03. This score was still higher than the *New York Daily News* Republican sentiment score of -0.11. *USA Today* had a Democratic sentiment score of 0.05 compared to a Republican sentiment score of -0.06 and the *LA Times* had Democratic and Republican sentiment scores of 0.11 and -0.05, respectively.

Table 4. Sentiment Scores at a Newspaper Level

Newspaper	Republican Tweets		Democratic Tweets		F-value	p-value
	N	Sentiment Score	N	Sentiment Score		
@denverpost	126	-0.02	62	-0.06	0.80	0.37
@ladailynews	91	0.01	36	0.01	0.02	0.87
@mercnews	150	-0.02	69	-0.07	0.38	0.54
@nhregister	13	-0.06	10	-0.15	0.22	0.64
@ocregister	17	0.07	12	0.14	0.001	0.98
@Enquirer	85	-0.03	39	-0.12	1.44	0.23
@USATODAY	597	-0.06	311	0.05	10.44	0.001***
@azcentral	93	-0.06	74	0.03	1.80	0.18
@indystar	77	0.02	24	-0.11	0.003	0.95
@journalsentinel	77	-0.04	37	0.10	1.99	0.16
@KCStar	135	-0.01	76	-0.03	<0.001	0.98
@MiamiHerald	145	-0.02	128	0.004	0.24	0.62
@startelegram	46	-0.07	27	0.01	0.13	0.72
@theobserver	57	-0.09	42	-0.002	0.77	0.38
@thestate	2	-0.42	1	-0.15	0.33	0.67
@NYDailyNews	509	-0.11	230	-0.03	5.84	0.02**
@SunSentinel	82	-0.11	64	0.02	2.23	0.14
@chicagotribune	289	-0.12	153	-0.12	0.22	0.63
@latimes	433	-0.05	242	0.11	9.83	0.002***
@sdut	13	-0.10	14	-0.10	0.001	0.97

* $p < 0.10$

** $p < 0.05$
*** $p < 0.01$

From these sentiment analysis results, one can conclude that some newspapers produced less negative tweets about the Democratic Party than the Republican Party leading up to the 2016 presidential election.

4.5 Tweets About Significant Events Leading Up to the Election

Two notable events that occurred in the month leading up to the election were the release of the Access Hollywood tapes of Donald Trump and Billy Bush discussing possible sexual misconduct and FBI Director James Comey announcing an investigation into new emails from Hillary Clinton's private email server. To examine tweets that covered these events, the Twitter data was searched for these keywords: "private server," "James Comey," "Comey," "FBI," "emails," "Clinton," "Access Hollywood," "Billy Bush," "Trump," "tape," and "AccessHollywood." The last keyword is included to catch hashtags mentioning Access Hollywood without tagging tweets with only "access" or "Hollywood" in them. Tweets were flagged if they contained two or more of these keywords, including hashtags.

Table 5. Tweets About Notable Election Events in the Election's Last 30 Days

Category	Number of Tweets
FBI Investigation into Clinton's Emails	306
Trump's Access Hollywood Tapes	303
Both Events Mentioned	3

Table 5 shows the number of tweets found referring to the two events — 306 tweets about the email investigation and 303 tweets about the Access Hollywood tapes. There were only three tweets that were categorized as related to both events. Figure 2 shows the number of tweets released per day referring to the two events.

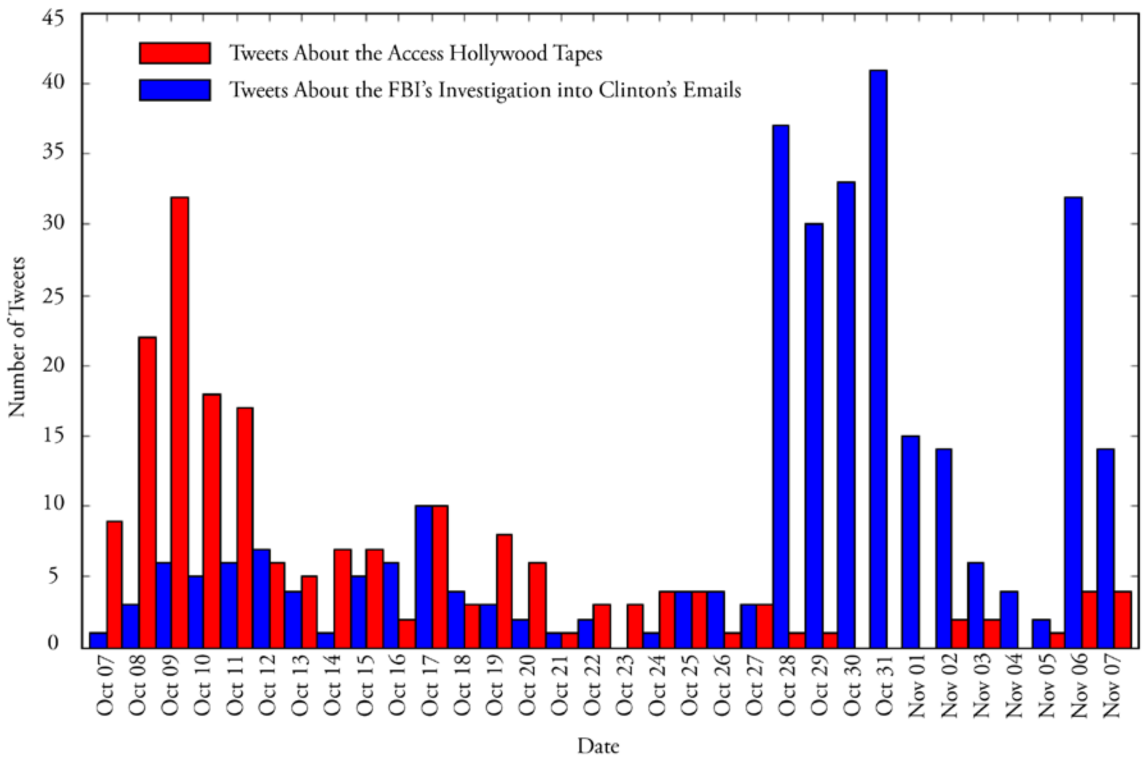

Figure 3. Tweets About Notable Election Events in the Election's Last 30 Days

The average sentiment score of the Clinton email tweets was a slightly negative -0.11 and the average sentiment score of the Trump tape tweets was -0.12. The ANOVA test between Clinton email tweets and Trump tape tweets resulted in an F-value of 0.12 and a p-value of 0.73. It is not surprising that the difference of means test was insignificant due to the similar average sentiment scores between groups. From this, one can conclude that the observed newspapers covered these two significant events during the last month of the presidential election similarly, in both coverage and tone because the number of tweets and sentiment scores were similar for both events.

An ANOVA test was also run for each newspaper to determine whether there was a significant difference between tweets about the FBI investigation into Clinton's emails and tweets about the Trump tapes at a newspaper level. The results of those ANOVA tests are in Table 6. Milwaukee's *Journal Sentinel* and Silicon Valley's *Mercury News* were the only two newspapers to produce significant ANOVA results, although these newspapers had a small number of tweets regarding the special events.

Table 6. ANOVA Results for Democratic and Republican Special Events Tweets

Newspaper	Tweets about FBI Investigation into Clinton's Emails		Tweets about Trump's Access Hollywood Tapes		F-value	p-value
	N	Sentiment Score	N	Sentiment Score		
@denverpost	16	-0.17	3	-0.14	0.03	0.86
@ladailynews	2	-0.26	4	0.10	0.71	0.45
@mercnews	9	-0.23	6	0.00	4.02	0.07*
@nhregister	3	-0.10	1	0.00	0.25	0.67
@ocregister	1	0.10	1	0.69	-	-
@Enquirer	5	-0.14	3	0.01	0.45	0.52
@USATODAY	57	0.02	40	-0.08	2.40	0.12
@azcentral	3	-0.09	3	-0.13	0.08	0.79
@indystar	5	0.00	3	0.38	1.92	0.25
@journalsentinel	5	-0.05	3	-0.55	5.69	0.05*
@KCStar	13	-0.14	6	-0.04	0.44	0.51
@MiamiHerald	15	-0.09	7	-0.14	0.05	0.825
@startelegram	8	-0.07	-	-	-	-
@theobserver	7	0.01	20	-0.13	0.71	0.41
@thestate	-		-		-	-
@NYDailyNews	58	-0.14	34	-0.16	0.05	0.82
@SunSentinel	10	-0.15	3	-0.41	2.75	0.13
@chicagotribune	49	-0.24	18	-0.20	0.16	0.79
@latimes	38	-0.04	29	-0.13	1.07	0.30
@sdut	5	-0.06	2	0.00	0.08	0.78

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

The newspaper with the highest number of Clinton email and Trump tape tweets was *USA Today*, with 40 and 57 tweets respectively. However, there was not a statistically significant difference between the sentiment scores for the two categories of tweets. *USA*

Today had a sentiment score of 0.02 for Clinton email tweets and a -0.08 sentiment score for Trump tape tweets. The analysis of these significant events at a newspaper level still supports the conclusion that newspapers covered the events similarly.

5 Conclusion

This exploration of employing automated sentiment analysis to study media bias in the last part of the 2016 presidential election show that media companies may not drive bias in the media. Moreover, there are many opportunities to expand the research on this application of sentiment analysis.

The analyses show that it is more likely to see a difference in sentiment towards political parties at a newspaper or local level than at a media company level. The sentiment analyses showed that in the month leading up to the 2016 presidential election, the *LA Times*, *USA Toda*y, and the *New York Daily News* produced tweets that were slightly more positive towards Hillary Clinton and the Democratic Party than their tweets about Donald Trump and the Republican Party. Additionally, the analyses showed that there was no significant difference in how newspapers covered James Comey's announcements on the FBI investigation into Hillary Clinton's private email server and the surfacing of the Access Hollywood tapes that recorded Donald Trump talking of sexual exploitation.

From these results a news consumer can conclude that it may be more important to diversify one's news intake across multiple newspapers rather than across multiple media companies in order to get balanced news, including following multiple news Twitter accounts. Researchers can conclude that there is room to explore how automated sentiment analysis can be applied to new corpuses.

Small sample sizes and the lack of a custom trained sentiment analysis model limit the results of this study. Due to the small sample sizes of some tweet categories, the analyses in this paper may not be generalizable to other time periods of interest or to other newspapers. Additionally, the out-of-the-box sentiment model was not trained on tweets specifically, so this may have affected the output. Both of these limitations can be addressed through future research.

Future research should strive for larger sample sizes to improve external validity. This could mean gathering tweets from more newspapers over a larger period of time or looking at only the newspapers that are most active on Twitter. It could mean using a different corpus, such as a dataset of newspaper headlines instead of Twitter.

Future research could also use a sentiment model that was trained to classify tweets including using Twitter data such as emoticon information and hashtags as classification features. Furthermore, a machine learning model interpreter such as Local Interpretable Model-Agnostic Explanations (LIME) could be implemented to better understand how the model is classifying sentiment. An interpreter would show the researcher what snippets of text, emoticons, or hashtags in an example led to the resulting sentiment score.

6 References

Chong, Wie Yen, Bhawani Selvaretnam and Lay-Ki Soon. "Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets." Paper presented at the *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*, Kota Kinabalu, Sabah, Malaysia, December 3-5, 2014.

D'Alessio, Dave and Mike Allen. "Media Bias in Presidential Elections: A Meta-Analysis." *Journal of Communication* 50, no. 4 (2000): 133–156.

Eberl, Jakob-Moritz, Hajo G. Boomgaarden, and Markus Wagner. "One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences." *Communication Research* 44, no. 8 (2017): 1125–1148.

Gallup. "Six in 10 in U.S. See Partisan Bias in News Media." http://news.gallup.com/poll/207794/six-partisan-bias-news-media.aspx (accessed April 13, 2018).

Gentzkow, Matthew, Edward L. Glaeser and Claudia Goldin. "The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered." *National Bureau of Economic Research Corruption and Reform: Lessons from America's Economic History* (2006).

Go, Alec, Richa Bhayani, and Lei Huang. "Twitter Sentiment Classification Using Distant Supervision." *CS224N Project Report, Stanford* 1.12 (2009).

Groseclose, Tim and Jeffrey Milyo. "A Measure of Media Bias." *The Quarterly Journal of Economics* 120, no. 4 (2005): 1191–1237.

Haselmayer, Martin, and Marcelo Jenny. "Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding." *Quality & Quantity* 51, no. 6 (2017): 2623-2646.

Hutto, C.J. and E.E. Gilbert. “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” *Eighth International Conference on Weblogs and Social Media* (2014).

Lott, John R. and Kevin A. Hassett. “Is Newspaper Coverage of Economic Events Politically Biased?” *Public Choice* 160, no. 1-2 (2004): 65–108.

“Number of Monthly Active Twitter Users in the United States From 1st Quarter 2010 to 4th Quarter 2017 (In Millions).” *Statista*. Last modified February 2018. https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/.

“Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 4th Quarter 2017 (In Millions).” *Statista*. Last modified February 2018. https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." *International Conference on Web and Social Media* 11 (2010): 122-129.

Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." Paper presented at the International Conference on Language Resources and Evaluation, Valletta, Malta, May 17-23 2010.

Pew Research Center. “Political Polarization and Media Habits.”

http://www.journalism.org/2014/10/21/political-polarization-media-habits/ (accessed April 14, 2018).

Santorini, Beatrice. “Part-of-Speech Tagging Guidelines for the Penn Treebank Project.” *Linguistic Data Consortium* (1990).

Soroka, Stuart. "The Gatekeeping Function: Distributions of Information in Media and the Real World." *Journal Of Politics* 74, no. 2 (2012): 514-528.

Soroka, Stuart and Lori Young. “Affective News: The Automated Coding of Sentiment in Political Texts.” *Political Communication* 29 (2012): 205–23.

Waykar, Pranav, Kailash Wadhwani, and Pooja More. "Sentiment Analysis in Twitter Using Natural Language Processing (NLP) and Classification Algorithm." *International Journal of Advanced Research in Computer Engineering & Technology* 5, no. 1 (2016): 79-81.

7 Curriculum Vita

Taylor Thomsen was born in Fort Riley, Kansas in 1993. She graduated from the University of Nebraska-Lincoln in 2015 with a Bachelor of Arts in Psychology and minors in Mathematics and National Security Studies. She earned her Master of Science degree in Government Analytics from Johns Hopkins University in 2018.