SYM-AM-20-061

PROCEEDINGS

OF THE

SEVENTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

**Acquisition Research:
Creating Synergy for Informed Change**

May 13–14, 2020

Published: April 13, 2020

ACQUISITION RESEARCH PROGRAM:
CREATING SYNERGY FOR INFORMED CHANGE

ACQUISITION RESEARCH PROGRAM:
CREATING SYNERGY FOR INFORMED CHANGE

Using Natural Language Processing, Sentiment Analysis, and Text Mining to Determine if Text in Selected Acquisition Report Executive Summaries Are Highly Correlated with Major Defense Acquisition Program (MDAP) Unit Costs and Can Be Used as a Variable to Predict Future MDAP Costs

Brian B. Joseph—is a Senior Operations Research Analyst in the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD[A&S]), Acquisition Analytics and Policy (AAP). In this role, Joseph serves as the Deputy Director of Data Analytics, supporting the Deputy to the Assistant Secretary of Defense, Acquisition Enablers, regarding acquisition data, data science, metrics, and statistics. Prior to this, Joseph served as the Statistician in the Security Policy & Oversight Division (SPOD), Counterintelligence and Security Directorate, Office of the Under Secretary of Defense for Intelligence. In this role, Joseph served as Principal Action Officer to the Director regarding DoD personnel security investigations metrics. Joseph earned an MS in cost estimation and analysis from the Naval Postgraduate School, an MS in applied statistics from the Rochester Institute, a BS in environmental management from the University of Maryland University College, and an AS in mathematics from Northern Virginia Community College. He is currently pursuing a PhD in data science from Northcentral University. [Brian.b.joseph.civ@mail.mil]

Darris Sconion—is a Data Analyst in the Acquisition, Analytics, and Policy Directorate of OUSD(A&S), analyzing DoD acquisition data. He earned his BS from Frostburg State University and is completing his MPA at Bowie State University. [Darris.a.sconion.ctr@mail.mil]

Other co-contributors from AAP: Michael Smith

Abstract

Major Defense Acquisition Programs (MDAPs) are required to report cost, performance, and schedule information updates to Congress annually via a Selected Acquisition Report (SAR). One of the components of the SAR is its executive summary, which provides an updated outlook of the health of the MDAP as well as what direction performance metrics may be trending. The executive summary is entirely textual. Traditional MDAP analysis is conducted using structured, continuous, and categorical data attributes. However, analysis of text to predict program metrics has rarely been used. This research conducts sentiment analysis of SAR executive summaries to determine whether their average emotional valence sentiment is highly correlated with MDAP unit cost metrics. Negative correlation depicts that, as average emotional valence sentiment increases, unit cost decreases, and positive correlation depicts that as average sentiment increases, so does its unit cost. If the results show high correlation, then average sentiment in the SAR executive summary may possibly be used as a primary or proxy variable in models that predict future MDAP costs. The results of our study found that, at most, only 12% of MDAP SAR executive summaries produce strong correlations ($|r| >= 0.70$) to possibly predict future MDAP costs.

Key Words

Correlation, SAR executive summary, MDAP, natural language processing, SAR, prediction, sentiment analysis, text mining

Research Issue/Business Need

The Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD[A&S]) has been examining ways to utilize the abundance of unstructured text data available in databases such as the Acquisition Information Repository (AIR) and the

Defense Acquisition Management Information Retrieval (DAMIR). Additionally, Fiscal Year (FY) 2017 to FY2019 National Defense Authorization Acts (NDAA) have urged the Department of Defense (DoD) acquisition community to use analytics to improve acquisition outcomes. SAR summaries are one of the most available and unutilized sources of textual data for text and sentiment analysis. The OUSD(A&S) wants to determine whether SAR text is useful in helping to create prediction models for MDAP costs.

Not many text analysis efforts have been conducted in the defense acquisition field. In fact, most of the recent work has been conducted by the Air Force Institute of Technology. McGowin, Ritschel, Fass, and Boehmke (2018) utilized text analysis to examine gaps in legislative enactments as they compared to a compendium of acquisition experts. Brown (2017) used text analysis of NDAAs to determine the relevance of cost estimating over a 20-year period and hypothesized that the frequency of cost estimating terms should appear more frequently over time. Freeman (2013) used Naïve Bayes supervised learning text classification of Format 5 text from Defense Cost and Resource Centers to predict program cost growth. Finally, Miller (2012). The results of this research will advance the use of text mining as a viable analysis method to assist in possibly predicting future MDAP costs.

Research Question

Are texts from SAR executive summaries highly correlated with MDAP unit cost metrics?

Hypothesis

H_0: At least 60% of average emotional valence sentiment in SAR executive summaries are correlated with unit cost metrics by $|r|>=0.70$

H_a: At least 60% of average emotional valence sentiment in SAR executive summaries are not correlated with unit cost metrics by $|r|>=0.70$

Related Work

There has been an emerging trend where financial market researchers have hypothesized that sentiment from Twitter is correlated with stock market trends and can be used to predict future stock market trends. Their ultimate goal is to show that the efficient market hypothesis (EMH) can be refuted. EMH states that it is impossible to beat the stock market consistently in the long run because stock prices fully reflect all of the information about the market.

Lansing and Tubbs (2018) used classical momentum theory in conjunction with sentiment analysis to predict stock returns of the Standard and Poor's index. The conclusion of this research was that the predictive power of the model increases when overall sentiment of the stock is declining and its momentum is in a negative state over a one-year period. Lansing and Tubbs found that during these periods there was increased investor interest, and investors begin to sell off the stock.

Ranco, Aleksovski, Caldarelli, Grcar, and Mozetic (2015) used sentiment analysis to examine the polarity of high volume tweets and their dependence on abnormal returns. They found that there is a positive correlation sentiment polarity of Twitter peaks implies the direction of cumulative abnormal returns.

Pagolu, Challa, Panda, and Majhi (2016) applied sentiment analysis and supervised machine learning principles to the tweets extracted from Twitter and analyzed the correlation between stock market movements of a company and sentiments in tweets.

Methodology

The methods for conducting the analysis are to (1) collect SAR executive summaries from actively reporting MDAPs in the DAMIR database, (2) create a corpus for each MDAP SAR executive summary, (3) clean the corpus to remove undesired text, (4) create negative/positive word clouds of the corpuses (5) extract sentiment features and metrics from each corpus and record average sentiment by year, (6) extract unit cost information from DAMIR for each MDAP by year, (7) conduct exploratory data analysis, and (8) conduct correlation analysis. Analyses for this study will be conducted using R programming language.

Data Collection and Preprocessing

We extracted 980 SAR executive summary texts from 86 actively reporting MDAPs (31 Air Force, 18 Army, and 37 Navy) from the DAMIR database. We also collected unit cost metrics for these MDAPs. Next we ingested the text into R programming language and created corpuses of each executive summary using the TM package in R by using a few techniques.

1. We tokenized the text to create individual words.
2. We removed English stop words that don't convey any emotion, such as *is* or *etc.*
3. All capital letters were transformed to lowercase to insure that lowercase and uppercase words that are the same were not duplicated.
4. Whitespace, punctuations, and numbers were removed.

Word Clouds

Word clouds allow for the analyst to obtain a visual representation of the most frequent words in the text (Silge & Robinson, 2017). We created word clouds for each MDAP and each year that the program reported any type of SAR (annual, exception, etc.). A total of 980 word clouds were created. We also created negative and positive word clouds for all SAR executive summaries to visualize the level of positive and negative sentiment in the text. Figures 1 and 2 below illustrate the two word cloud versions for the F-35 aircraft, created from the 1997 SAR executive summary.

Figure 1.   Word Cloud of F-35 1997 SAR Executive Summary

Figure 2.   Positive/Negative Word Cloud of F-35 1997 SAR Executive Summary

Sentiment Analysis

Text mining is the process of deriving high-quality information from text. One of its applications is sentiment analysis, where the polarity of text is analyzed to determine whether its content is positive or negative (Silge & Robinson 2017). Emotional valence and emotional propensity are two metrics used to measure the emotion of words in text. Emotional valence measures the polarity of the text on a −1 to +1 scale with negative values mapping to negative words in the text and positive values mapping to positive words in the text (Awesome Open Source, n.d.). Zero values equate to neutral text. Emotional valence occurs over the length of the text. In contrast, emotional propensity measures the presence of eight psychological factors that may be contained in text. The factors include anger, fear, anticipation, disgust, joy, sadness, surprise, and trust. The impetus for this study is based on the analysis of emotional valence; however, we present illustrations of both emotion metrics in sentiment analysis. We utilized the sentimentr package in R programming language to extract the average sentiment over the length of the text for each year per each SAR executive summary (Rinker, 2019). Figures 3 and 4 are plots of emotional valence and emotional propensity for the F-35 in 1997. Figure 3 illustrates that the emotional valence for the 1997 SAR executive summary increases from −0.5 to 1 for the first quarter of the document, decreases from 1 to −0.5 for the second quarter of the document, increases from −0.5 to 0.0 for the third quarter of the document, and finally decreases from 0.0 to −1.0 over the final quarter of the document. The average emotional valence for the 1997 F-35 SAR executive summary is 0.16.
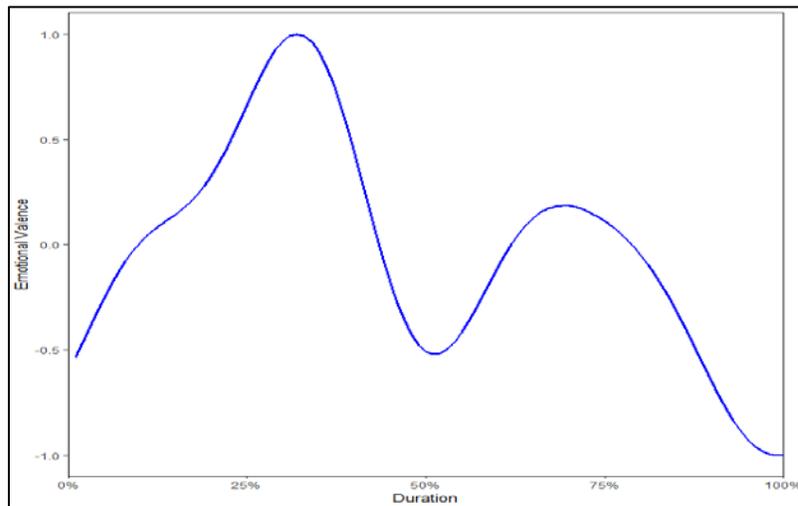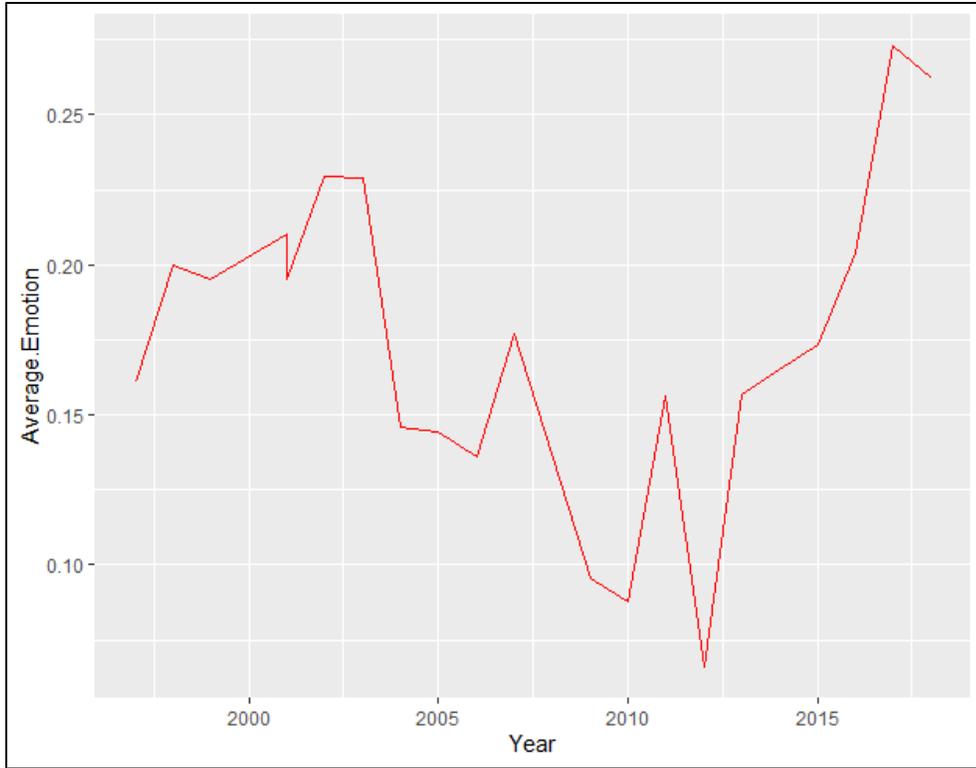

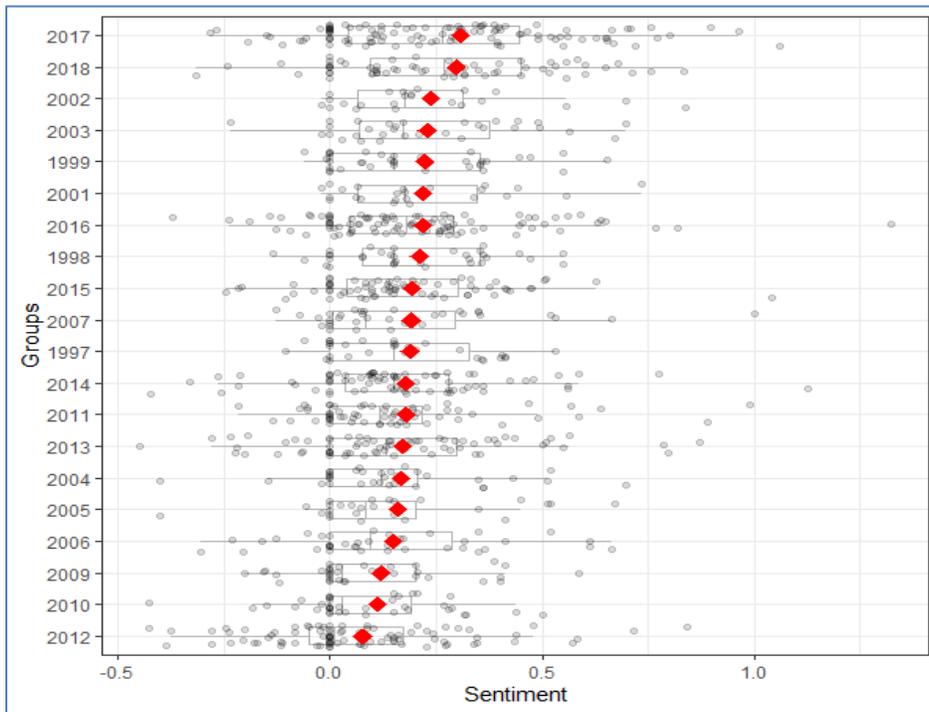**Figure 3. Emotional Valence of F-35 1997 SAR Executive Summary**
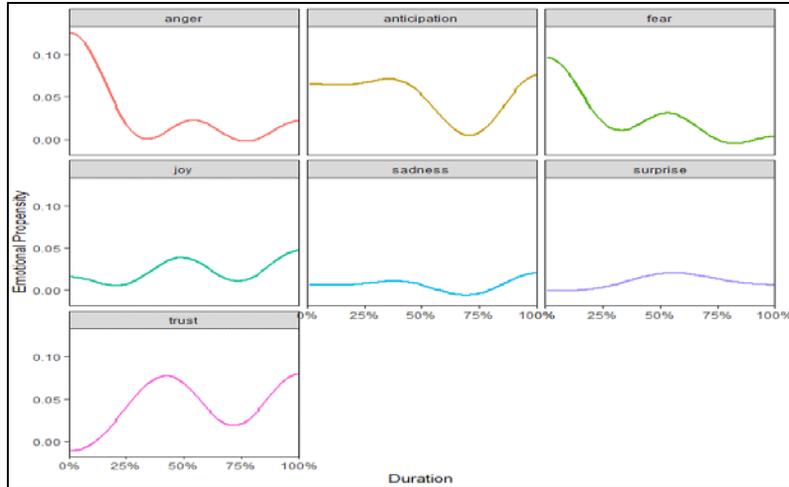
Figure 4.   **Emotional Valence of F-35 1997 SAR Executive Summary**

Correlation of Average Emotion Sentiment versus APUC and PAUC

Figure 5 illustrates the average emotional valence for the F-35 aircraft from 1997 to 2018. Figure 6 illustrates the boxplot average emotional valence sorted from largest to smallest by year. You can see that 2012 had the lowest average emotional valence sentiment while 2017 had the highest. Table 1 displays the overall average emotional valence statistics for the F-35 from 1997-2018. We also calculated average emotional valence for each of the remaining 85 MDAPs in our research sample. We used the average emotional valence sentiment metric from the SAR executive summaries in conjunction with current and original estimates of percentage increases/decreases in average procurement unit costs (APUC) and program acquisition unit costs (PAUC) of MDAPs to determine if average emotional valence is correlated with these unit cost metrics and could ultimately be used to predict future MDAP costs. Our logic was based on related research that used information from tweets to predict stock market returns. Our hopes are that the changes in average emotional valence correlate with percent changes in unit cost. At the end of the correlation analysis, the number of MDAPs were culled to 69 due to sparse numbers of SAR executive summaries and unit cost entries in DAMIR.

Figure 5. Average Emotional Valence of F-35 SAR Executive Summaries from 1997 to 2018

Figure 6. Sorted Boxplots of F-35 Average Emotional Valence 1997-2018

**Table 1. Average Emotional Valence
Sentiment of F-35 SAR Executive
Summaries**

Year	Word Count	SD	Average Emotional Valence Sentiment
1997	586	0.17	0.19
1998	671	0.18	0.21
1999	724	0.19	0.22
2001	507	0.19	0.22
2002	501	0.22	0.24
2003	655	0.22	0.23
2004	597	0.21	0.17
2005	502	0.23	0.16
2006	769	0.23	0.15
2007	721	0.23	0.19
2009	822	0.18	0.12
2010	981	0.18	0.11
2011	1291	0.22	0.18
2012	1713	0.23	0.08
2013	1596	0.25	0.17
2014	1909	0.24	0.18
2015	1631	0.21	0.19
2016	2494	0.25	0.22
2017	2772	0.28	0.31
2018	1551	0.25	0.30

We used R programming language to calculate the correlation (r) between the average emotional valence of the SAR executive summary and the percent increase/decrease of the four unit cost metrics (current and original estimates for APUC and PAUC) over the length of time each MDAP reported a SAR. Table 2 below illustrates the correlation results for the F-35 aircraft. Table 5 in appendix A displays the correlation results for all 69 MDAPs. The same analysis was conducted for the remaining 85 MDAPs. Table 3 shows that, at most, only 12% of current APUC and PAUC unit cost metrics and, at most, 13% of original APUC and PAUC metrics were correlated with average emotional valence sentiment in SAR executive summaries.

Table 2. F-35 Correlation Average Emotional Valence by Unit Cost Metrics

Program	% Change Current. PAUC	% Change Current APUC	% Change Original PAUC	% Change Original APUC	Commodity Type	Service
F-35 Combined	-0.834024	-0.8066895	-0.7771538	-0.7640996	Aircraft - Fighter	DoD
F-35 Aircraft	0.3850787	0.2628264	0.516665	0.4293511	Aircraft - Fighter	DoD
F-35 Engine	-0.2978912	-0.1186572	-0.1913041	-0.2814591	Aircraft - Fighter	DoD

Table 3. Proportion of Correlation Strength of Unit Cost Metrics (n=69 MDAPs)

Range	Strength of Association	% Change Current PAUC	% Change Current APUC	% Change Original PAUC	% Change Original APUC
0	No Association	0	0	0	0
>0-<0.25	Negligible	42.02	40.57	34.78	37.68
0.25-<0.50	Weak	24.63	30.43	26.08	27.43
0.50-<0.70	Moderate	21.73	17.39	26.08	21.73
0.70-<1	Strong	8.69	10.14	11.59	11.59
1	Perfect	2.88	1.44	1.44	1.44

Results

Based on the test of proportions in Table 3, we reject the null hypothesis and conclude that at least 60% of average emotional valence sentiment in SAR executive summaries are not correlated with unit cost metrics by |r|>=0.70.

Table 4. Proportion Test: At least 60% of average sentiment in SAR executive summaries are not correlated with unit cost metrics by |r|>=0.70 (n=69 MDAPs)

Metric	Proportion Estimate	LCL	UCL	p-value
% Change Current APUC	0.12	0.00	0.20	<0.0001
% Change Current PAUC	0.12	0.00	0.20	<0.0001
% Change Original APUC	0.13	0.00	0.22	<0.0001
% Change Original PAUC	0.13	0.00	0.22	<0.0001

Discussion and Conclusion

This paper aimed to use text analysis, natural language processing, and sentiment analysis to determine whether future MDAP cost could be predicted based on at least 60% of associations between average emotional valence sentiments in SAR executive summaries and unit cost metrics having a strong correlation ($|r| \geq 0.70$). Based on descriptive statistics and hypothesis testing of proportions, we found that, at most, only 12% of 69 MDAPs had strong correlation to predict future MDAP cost based on SAR executive summaries and unit cost metrics. As such, we should only pursue regression models for those MDAPs that achieved correlations of $|r| \geq 0.70$ based on average emotional valence as the independent variable and unit cost metrics as the dependent variable based on average emotional valence and unit cost metrics. We believe this study contributed to advancing the use of text mining and sentiment analysis in DoD acquisition to predict future MDAP costs using acquisition data from the DAVE/DAMIR database. The research demonstrated that SAR executive summary texts and unit cost information can possibly be utilized to predict future cost for 13% of MDAPs.

Limitations of Study

MDAPs produce relatively small sample sizes of SAR executive summaries across SAR reporting periods, and there are often not a one-to-one mapping to percent change unit cost metric mappings as some of that data is missing in the DAMIR database. To exacerbate the sample size issue, some programs are divided into subprograms, which makes it impossible to combine correlations, as we would be comparing metaphorical apples to oranges. Another limitation of this study is that the emotional valence lexicon for positive and negative words are not per se equivalent to an acquisition lexicon/dictionary. Therefore, words that may be negative in the typical English language may not be negative in the acquisition language. Finally, if there were a strong correlation between average emotional valence and unit cost metrics, and a valid model could be created to predict future MDAP costs, program offices may change the language they place into SAR executive summaries to salt the correlations and predictions.

Future Research

Based on these limitations, our future research will pursue machine learning techniques such as neural networks, support vector machines, and topological data analysis to predict future MDAP costs based on SAR executive summaries. To increase the sample sizes, we also propose to use Defense Acquisition Executive Summaries (DAES) to replicate this study as well as for future proposed machine learning studies. The DAES produce four times as many sample executive summaries as the SARs. Another approach would be to create an acquisition dictionary of negative and positive words and rerun the analysis. Finally, we should compare the results of regression models based on those 12% of MDAPs that had strong correlations in this study to machine learning models of raw SAR executive summary text and unit cost information to see which model performs best.

References

Brown, G. E. (2017). Measuring the increasing relevance of cost estimating through text analytics. *ICEAA World, 2017*(2), 32–33.

Freeman, C. E. (2013). *Multivariate and Naïve Bayes text classification approach to cost growth risk in Department of Defense acquisition programs* (Master's thesis, Air Force Institute of Technology). Retrieved from https://scholar.afit.edu/etd/953/

Lansing, K. J., & Tubbs, M. (2018). *Using sentiment and momentum to predict stock returns.* Retrieved from https://www.frbsf.org/economic-research/publications/economic-letter/2018/december/using-sentiment-and-momentum-to-predict-stock-returns/

McGowin, A. L., Ritschel, J. D., Fass, R. D., & Boehmke, B. C. (2018). A text mining analysis of acquisition reforms and expert views. *Defense Acquisition Research Journal*, *82*, 288–323.

Miller, T. P. (2012). *Acquisition program problem detection using text mining methods* (Master's thesis, Air Force Institute of Technology). Retrieved from https://scholar.afit.edu/etd/1022/

Mudinas, A., Zhang, D., & Levene, M. (2018). Market trend prediction using sentiment analysis: Lessons learned and paths forward. In *Proceedings of the 7th KDD Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '18).* ACM. Retrieved from https://arxiv.org/pdf/1903.05440.pdf

Pagolu, V. S., Challa, K. N. R., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System.* IEEE. Retrieved from https://ieeexplore.ieee.org/document/7955659

Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., & Mozetic, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE, 10*(9). Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0138441

Rinker, T. (2019). Package "sentimentr." Retrieved from https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf