

Sentiment Analysis and Video Assistant Referees in Premier League Soccer

Dylan Blechner

Department of Sport Analytics, Falk College of Sport and Human Dynamics, Syracuse University

Syracuse University



ABSTRACT

After the recent introduction of Video Assistant Referees (VAR) into professional soccer, newfound controversies have arisen regarding the nature of calls made during a given game. Referees and fans alike are allowed to use the instant replay technology to make their own judgements about on-field decisions. Many have taken to social media to voice their displeasure about the new technology. This raises the question: "Is it possible to determine whether or not a VAR call is 'correct' by using fan sentiment?" Through the use of ESPN commentary and Twitter data, this question can start to be answered. Through my research I hope to analyze VAR decisions in professional soccer to help improve its use in leagues and competitions. Because the use of VAR is quite subjective, I will show that analysis of fan sentiment can be used as a proxy to determine which calls may have been incorrect. There is a lot of work that can be done to improve VAR to make the refereeing of the game as fair and accurate as possible. A better use of this technology can improve television ratings in professional leagues and improve overall fan interest.

INTRODUCTION

Video Assistant Referees (VAR) are off-field match officials that review calls made or missed by the head on-field referee with the assistance of replay technology. VAR has been recently introduced to the game of soccer and has been used in the FIFA Men's and Women's World Cups, UEFA Champion's League, and each of the top five professional soccer leagues in Europe. In my analysis, I will test several hypotheses through the use of various data analyses. The first hypothesis is that VAR decisions during professional soccer matches increases discussion about VAR on social media. This may be a simple hypothesis, but if it is not met, then it is impossible to analyze sentiment associated with VAR as very few people would be talking about it. The second hypothesis is that the sentiment of tweets about VAR decisions is more negative rather than positive or neutral. I believe that while the use of VAR has improved the game of soccer, fans are not used to the types of calls that are made with the assistance of the technology. In my preliminary literature review, I found many non-academic articles that have discussed the use of VAR in various professional leagues and competitions. There has been considerable talk about the use of VAR in the English Premier League this season because this is the first year that the technology has been implemented. There were very few academic articles written about the use of VAR, but several papers have analyzed social media sentiment associated with soccer matches to quantify excitement related with specific game events. I wanted to perform a similar analysis with VAR decisions in professional soccer.

METHODS

Twitter data was gathered using the "twitterscraper" module in Python. Tweets containing the term "var" or the hashtag "#var" were scraped for each day that a Premier League game occurred during the first half of the 2019-2020 season as well as those containing "var" and a game-specific hashtag. Commentary data was also gathered for the EPL from ESPN by using the "fscrapR" package in R. This commentary data includes information about key events that took place during a given match including VAR decisions. Using these tweets, I was able to perform sentiment analysis to evaluate how fans felt about the new use of VAR in the Premier League. I first looked at how many tweets were posted each week to understand how fans felt about VAR over time. There were 158 VAR decisions made during the first half of the EPL season. There is a weak positive correlation between the number of VAR calls and the number of tweets about VAR during a given week meaning that we would expect fans to tweet more about VAR when there are more decisions that involve the technology.

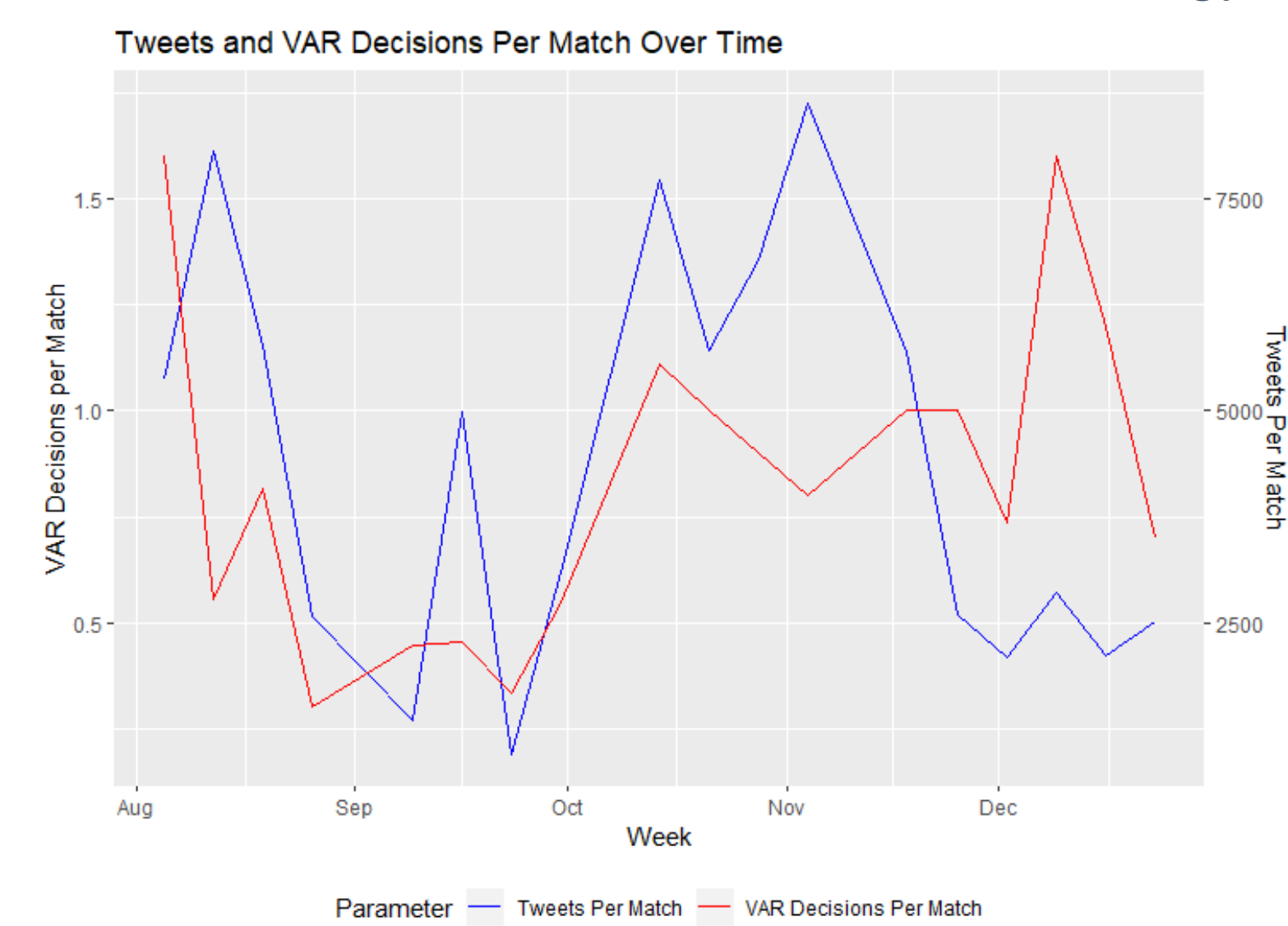


Figure 1: Number of tweets per match and number of VAR decisions per match

Sentiment can be broken down through the use of the NRC Lexicon which categorizes words into eight additional categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. There were 7,426 occurrences of negative terms followed by 6,697 positive terms which shows the negativity shown by tweeters towards VAR. One of the additional sentiment categories that appeared frequently was anger with 3,911 terms. A large portion of tweeters do not just have negative feelings towards the use of VAR, but are also angry with the decisions being made. Tweeters showed trust as well, demonstrating the divide amongst fans about the use of VAR.

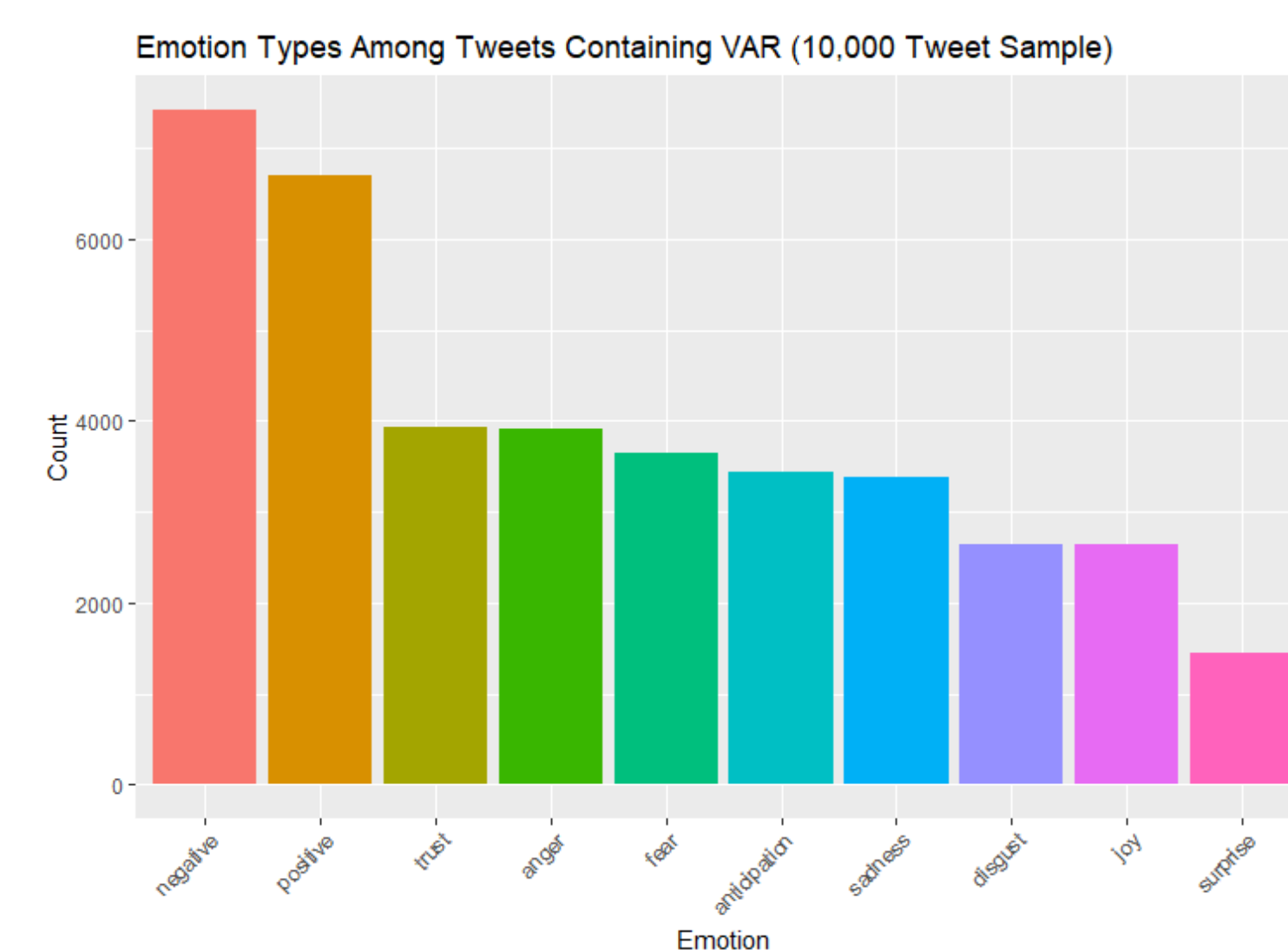


Figure 2: Frequencies of terms by NRC Lexicon Category

RESULTS

After analyzing fan sentiment on Twitter through the use of sentiment lexicons, I wanted to utilize Natural Language Processing techniques to create a model to predict the sentiment of future tweets concerning VAR. To train my model, I used Twitter data from the SemEval 2014 competition which consisted of approximately 10,000 tweets that were categorized as either positive, negative, or neutral. I ran a Naïve Bayes Classification model in Python with the NLTK package using feature sets to predict the sentiment of a given tweet. After tokenizing the tweets, I used the 1500 most common tokens as features in the model. I also utilized the Subjectivity, LIWC, and VADER sentiment lexicons to improve model performance.

Feature	Relationship	Feature	Relationship
***	neg : neu = 66.4 : 1	negativecount = 9	neg : neu = 19.5 : 1
:	pos : neg = 51.4 : 1	thanks	pos : neu = 19.4 : 1
fun	pos : neu = 49.5 : 1	missing	neg : pos = 18.3 : 1
excited	pos : neu = 47.2 : 1	negativecount = 10	neg : neu = 18.1 : 1
sad	neg : neu = 43.9 : 1	cool	pos : neu = 16.8 : 1
sorry	neg : neu = 43.9 : 1	cant	pos : neu = 15.1 : 1
positivecount = 12	pos : neu = 42.5 : 1	brilliant	pos : neu = 15.1 : 1
great	pos : neu = 38.9 : 1	funny	pos : neu = 14.2 : 1
happy	pos : neu = 35.7 : 1	exciting	pos : neu = 14.2 : 1
amazing	pos : neu = 35.7 : 1	wrong	neg : neu = 14.2 : 1
luck	pos : neu = 28.8 : 1	hate	neg : neu = 13.9 : 1
thank	pos : neu = 25.4 : 1	negativecount = 8	neg : pos = 13.8 : 1
:(neg : pos = 25.3 : 1	can't	neg : neu = 13.4 : 1
injury	neg : pos = 23.5 : 1	interesting	pos : neu = 13.3 : 1
awesome	pos : neu = 21.9 : 1	pavol	neg : pos = 13.1 : 1

Table 1: The 30 most informative features in the Naïve Bayes Classification model

I then used cross-validation to find precision, recall, and F1 scores. Using 10 folds, I found a mean accuracy of 0.687 or 68.7%. Overall, the model performed very well and can be used to predict the sentiment of the tweets about VAR in the English Premier League.

Sentiment Category	Precision	Recall	F1
Neutral	0.693	0.660	0.676
Negative	0.490	0.500	0.495
Positive	0.658	0.695	0.676

Table 2: Sentiment model accuracy scores

I then wanted to take a look at the breakdown of tweets during a match to see if there was more negative sentiment after VAR decisions. If we look at a match between Manchester City and West Ham, we can see that there were large spikes in neutral sentiment after each VAR decision. It can be seen that the first decision was split evenly in terms of positive and negative sentiment suggesting that fans were undecided on whether or not the decision was correct. The third call was met with more negative sentiment than positive. Fans were not pleased with the third call because of the nature of the decision which led to the Manchester City penalty being retaken for a minor infringement. This information can be very useful to the Premier League as they continue to improve the use of VAR.

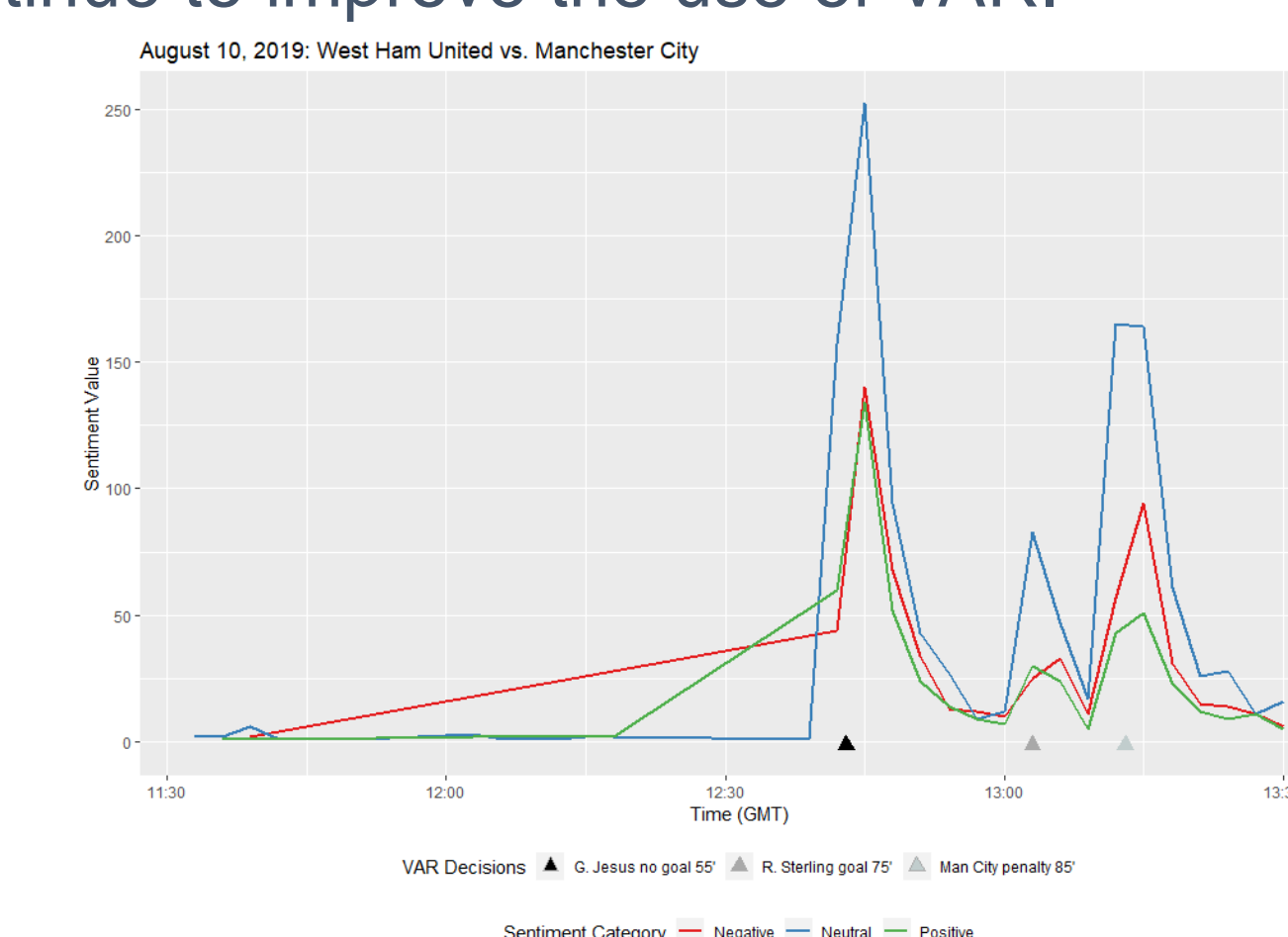


Figure 3: Predicted sentiment during the Manchester City vs. West Ham match

CONCLUSION

In conclusion, I have utilized Twitter data to analyze fan sentiment associated with Video Assistant Referee decisions in professional soccer. I first found that VAR decisions during professional soccer matches does in fact increase social media activity. As the number of VAR decisions in a match increases, we would generally expect the number of tweets about VAR to increase as well. I also found that the sentiment of tweets about VAR decisions is more negative rather than positive or neutral. Through the use of the Bing and NRC Lexicons it is clear that fans show more negativity towards the use of VAR than positivity or indifference. I was also successful in creating a model to predict the sentiment associated with a given tweet concerning VAR through the use of Natural Language Processing. The model helped to emphasize the negative nature of tweets about VAR in the English Premier League. Through my analysis, changes to VAR can be implemented to further improve the game. VAR technology has already improved the game of soccer, and by repairing the flaws in the system, it will continue to help the beautiful game.

REFERENCES

- Goff, S. (2019, June 26). Analysis | VAR is working at World Cup, FIFA says. That's not wrong, but it could work better. Retrieved September 10, 2019, from <https://www.washingtonpost.com/sports/2019/06/26/var-is-working-world-cup-fifa-says-thats-not-wrong-it-could-work-better/?hpid=hp-sports%3Aworld-cup%3Ahomepage%2Fstory&hpid=hp-sports%3Aworld-cup%3Ahomepage%2Fstory>.
- Johnson, D. (2019, August 10). VAR in the Premier League: The big decisions explained. Retrieved August 13, 2019, from <https://www.espn.com/soccer/english-premierleague/story/3917260/var-in-the-premier-league-the-big-decisions-explained>.
- VAR in the Premier League: How did first weekend go for technology in top flight? (2019, August 11). Retrieved August 13, 2019, from <https://www.bbc.com/sport/football/49307496>.
- Yu, Y., & Wang, X. (2015, February 20). World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets. Retrieved October 3, 2019, from <https://www.sciencedirect.com/science/article/pii/S074756321500103X>.
- Corney, D., Martin, C., & Goker, A. (2014, April 13). Spot the Ball: Detecting Sports Events on Twitter. Retrieved October 3, 2019, from http://dcorney.com/papers/ECIR_SpotTheBall.pdf.
- Gratch, J., Lucas, G., Malandrakis, N., Szablowski, E., Fessler, E., & Nichols, J. (2015, September 21). GOAALLL! Using Sentiment in the World Cup to Explore Theories of Emotion. Retrieved October 5, 2019, from <http://ict.usc.edu/pubs/GOAALLL!UsingSentimentintheWorldCuptoExploreTheoriesofEmotion.pdf>.
- Taspinar, A. (2019, November 4). TwitterScraper. Retrieved December 20, 2019, from <https://github.com/taspinar/twitterscraper>.
- Yurko, R. (2018, July 5). FScrapR. Retrieved August 1, 2019, from <https://github.com/ryurko/fscrapR>.
- Wasser, L., & Farmer, C. (2017, April 19). Sentiment Analysis of Colorado Flood Tweets in R. Retrieved January 8, 2020, from <https://www.earthdatascience.org/courses/earthanalytics/get-data-using-apis/sentiment-analysis-of-twitter-data-r/>.
- Mejova, Y., Srinivasan, P., & Boynton, B. (2013, February 1). GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media. Retrieved January 14, 2020, from <https://dl.acm.org/doi/10.1145/2433396.2433463>.
- Beal, V. (n.d.). Huge List of Texting and Online Chat Abbreviations. Retrieved September 15, 2019, from <https://www.webopedia.com>.
- TweetTokenizer. (n.d.). Retrieved September 15, 2019, from <https://kite.com/python/docs/nltk.TweetTokenizer>.
- Hutto, C. J. (2019, September 6). VADER Sentiment Analysis. Retrieved September 15, 2019, from <https://github.com/cjhutto/vaderSentiment>.

ACKNOWLEDGEMENTS

I'd like to say thanks to my faculty advisor Dr. Rodney Paul, who helped me through every step of the writing process of my undergraduate thesis. I'd also like to thank Falk College for funding my trip to Atlanta where I presented and won the Undergraduate Paper Competition at the AEF Conference.