Operation Tweetstorm: the influence of information shared on twitter

Jeffrey Lee jlee3703@gatech.edu
Minhao Leong mleong7@gatech.edu
Jinchao Lin jlin453@gatech.edu
Jun Xiong Tan jtan318@gatech.edu
Dr Matthew Tinker mtinker6@gatech.edu

1. Introduction

Since the introduction of social media, the spread of news and information has significantly fastened (Westerman, Spence, Van Der Heide, 2014). Although social media is not new, the digital channel on which it exists has enabled faster spread of material (Lerman, Ghosh, 2010) and with increased access, people can choose which information to consume (Karlsen, Steen-Johnsen, Wollebæk, Enjolras, 2017). This report will outline how Team 26 has developed a platform to understand the influence that twitter has on world events and the lessons that have been learnt in developing a progressive approach to this problem.

2. Survey

Twitter, "a popular microblogging service where users create messages (called "tweets")" (Go, Bhayani, Huang, 2009) offers access to services to analyze tweets. Social media's influence is increasingly relevant and important to understand (Allcott, Gentzkow, 2017). Various research has been conducted including a finding that public sentiment and voting preferences can be identified by tweet volume (Tumasjan, Sprenger, Sandner, Welpe, 2010). Reviewing methods currently used there is both value and shortcomings of each.

Himelboim, Smith, Rainie, Schneiderman, and Espina (2017) looked at network structures on Twitter to categorize the communities that exist. It did not categorize the tweets and will not be used in the first iteration of our development. Research on the extraction of topical keyphrases found that context-sensitive topical page rank observing relevance and interestingness boosted performance of keyphrase extraction (Zhao et al. 2011). Whilst more relevant to real-time posts it may assist in the data extraction process in later work.

In research linking tweets to poll results, O'Connor, Balasubramanyan, Routledge and Smith (2010) found that simple sentiment analysis can replicate presidential approval and consumer confidence. This shows that tweet sentiment could be linked to the outcome of events. The analysis itself used simple text analysis on initial tweets and could be expanded to retweets, forwards and analysis beyond text. Further to this, Tumasjan, Sprenger, Sandner, and Welpe (2010) found that the number of tweets about politicians in the German federal election could be predictive of voter preferences and sentiment provided insight of the campaign messages. The approach was reflected upon for Team 26's project however they were only able to look at tweets containing politician and party names whilst our research has expanded beyond this. The outcome of this research was also disputed due to "arbitrary choices of the authors" (Jungherr, Jurgens, Schoen, 2011, p. 229).

The bias of Twitter account users through political leaning scores was reviewed by Wong, Tan, Sen, and Chiang (2013) but they didn't consider user following and retweet sentiment. This research could add an interesting dynamic in our model although this would be a later extension. Further, Karlsen et al. (2017) found that whilst users seek debate, it doesn't serve to change people's minds but rather reinforces their opinions.

Sentiment is important to the proposed model and a real-time model for the 2012 US election found that tweet volume was driven by campaign activity (Wang, Can, Kazemzadeh, Bar, and Narayana, 2012), we conclude that this may suggest undue influence in the moment and so we instead undertook a retrospective review. Giachanou and Crestani (2016) also looked at sentiment, reviewing over 50 papers to look at the "non-trivial task" (p. 6) of sentiment analysis. In building our model, we considered all factors they've outlined including identifying opinion, irony, and emotion.

One model, proposed by Pandey, Rajpoot and Saraswat (2017) who used a combination of techniques including cuckoo search and K-means in order to understand tweet sentiment was considered for our approach. The method was comprehensive including classifying emoji's. Irony, sarcasm and the length of tweets was not handled in the model although with knowledge of the gaps, further work could close these. Wilson, Wiebe and Hoffman (2009) took sentiment analysis and added polarity measures to the processing of text, which is an important feature for our project. This work did not look at symbols and jargon that we see on twitter.

Use of an existing lexicon will greatly speed up our work and Hasam, Moin, Karim and

Shamshirband (2018) found TextBlob or W-WSD sentiment lexicons perform best. A challenge not addressed by this but by Jianqiang and Xiaolin (2017) is the use of acronyms, hyperlinks, and negating words who found a method that increased the accuracy of sentiment classification when negation is replaced, and acronyms are expanded. Text pre-processing was a critical step of our research and this was a key focus in our data preparation.

An area that needs consideration is the handling of symbols. Emoji smoothed language models were proposed by Liu, Li and Guo (2012) where they utilized both manually and noisy labelled data in their sentiment analysis however the analysis is limiting at scale and was not used. Emoji use was also explored by Rakhmetullina, Trautmann and Groh (2018), Wolny (2016) and Wood and Ruder (2016) who found that they may be useful as emotion labels for sentiment, although their approach was also manual. Sari, Ratnasari, Mutrofin, and Arifin (2014) developed a method that addresses these issues which has subsequently been factored into our approach.

Van Hee, Lefever, and Hoste (2018) provided a model to approach irony in tweets which was to use a series of hashtags. Whilst this approach is limiting it assisted our project with initial work on the detection of sarcasm in our analysis, this was not included in our final model, however. Further challenge exists in irony with emoji use and Chen, Yuan, You and Luo (2018) created an approach that considers words and the emoji used to detect the sentiment. Their model was a good advance on standard text-based approaches and was factored into our work.

Following review of the research, it was proposed that a platform be developed for users to visualize the influence of twitter on event outcomes. More use of social media means we need a way to see how it's influencing major events.

3. Problem

The motivation of this work was to understand social media's influence on the outcome of events given its growing influence on the content people consume.

The problem that has been addressed by this piece of work is the development of understanding of how tweet sentiment can be used as a predictor on event outcomes.

Whilst we have used the 2016 US Presidential Election to test our model and platform, we believe that our platform can be easily adapted to any event whereby there is two potential outcomes particularly a winner or loser outcome.

4. State of the art approach

The platform that has been developed expands upon the works undertaken previously by others and uses a combination of the techniques including complex system of rules (including consideration of negation and intensity) for the development of our sentiment analysis, along with the data preparation steps proposed by Jianqiang and Xiaolin (2017) together with the work on emoji handling proposed by Sari, Ratnasari, Mutrofin, and Arifin (2014). The innovation in this work is the combination of the previous works together with a user-friendly platform for users to review the model outputs.

5. Method to address the problem

The method undertaken to address the problem, as it was defined, included extracting 10 million tweets from the days of five key milestones, including the debates, that occurred in the lead-up to the 2016 US Presidential election. Each of the tweets extracted was cleansed to remove duplication of words and irrelevant characters.

The data extracted was stored in an SQLite database which we then used to prepare a series of queries for the use of our Tweetstorm platform.

5.1 Data Preparation

Over 10 million tweets from key stages of the 2016 US Presidential Election were obtained by using the ID's captured by the Harvard Dataverse (Littman, Wrubel, Kerchner, 2016) and then using these tweet ID's the team called the Twitter API to hydrate the IDs to obtain the full tweets. Once these tweets were extracted, they were run through a series of cleansing processes which were conducted to remove tags, links, and non-useful labels including "RT" for retweet. These steps then provided a clean and useful dataset to allow us to progress to the next stage.

The further expansion upon this data preparation was to commence classifying characters produced as a result of the use of emoji's. Characters used in combination that could be classified as emoji's were left in the dataset for the use of the analysis. These tweets were store on an SQLite Database.

5.2 Data Queries

Because of the historical nature of the data, once there was a clean set of tweets available, the team decided that to improve the performance of the platform, a set of queries would be run to produce static datasets. The platform will be capable of running real-time queries in further iterations but was not required in this instance.

The queries that were developed focused on the information required for the purpose of conducting the sentiment and polarity models and then subsequently for the purposes of undertaking the visualizations. The queries included geographic mapping of the source of the tweets, sentiment and polarity of the tweets, and volume of the tweets.

5.3 Sentiment Model

Following the preparation of the data the focus was on developing the model for sentiment and polarity. To conduct this piece of work the team produced a sentiment analysis tool using a complex system of coded rules. This approach was taken in order to capture intricacies such as negation and differing intensities implied by different words. To ensure optimal performance of the model, significant text pre-processing was applied by removing invalid characters using regex, as well as lemmatization of words based on their part-of-speech tagging.

The model then developed assigned both a polarity and a sentiment score to each tweet. The underlying approach works in the manner of a lookup table on a library of words and phrases which provided the polarity, subjectivity and intensity. For each word these scores are applied in the following way:

- Polarity – a score between -1 and 1, with the higher score representing a more positive sentiment;
- Subjectivity – a score between 0 and 1, 0 being objective, 1 being subjective;
- Intensity – a score which is a multiplier between 0.5 and 2 and considers a word and its influence on the following word.

There were some words where multiple metrics were generated, our model accounted for these by applying the average of the metrics.

In our model we realized that we had to make consideration for negation words including ‘not’ that would counter the word following, and this was accounted for through the scoring. The model was then run over the entire dataset of tweets and the scores for each tweet was appended to the original clean set of data.

5.4 Model Outcomes

Given the fact that the tweets do not come with a polarity and subjectivity score, the model could not be evaluated using the standard train and test set accuracy validation. This meant that innovative approaches were taken to address the unsupervised nature of the problem. Sentiments for the tweets had an expected distribution with most of the sentiments centered near the average as would be expected. Separate word clouds were also generated for the positive and negative polarity tweets, with meaningful words that supported the assigned sentiments. These indicators suggest a well-trained and calibrated sentiment model.

5.5 Visualization

The visualizations produced for Operation Tweetstorm use Javascript and the D3 library, the visuals include:

5.5.1 Geographic Polarity

The first key visualization was to create a geographic representation to allow the demonstration of the polarity and sentiment of tweets across the United States of America. The visualization allows the user to clearly see the polarity and subjectivity by State and at the 5 key milestones. Using these scores, it also shows the preference of each candidate at these points.

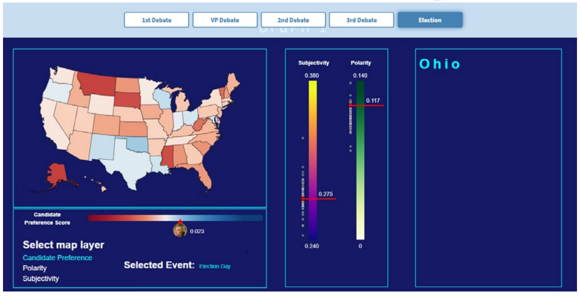


Figure 1: Geographic Sentiment

5.5.2 Polarity and Subjectivity Time Series

As an additional expansion upon the geographic visualization, the team prepared a view of the polarity and subjectivity over the five key time points during the election including the First Presidential Debate, the Vice-Presidential Debate, the Second Presidential Debate, the Third Presidential Debate, and Election Day. This clearly shows the movements through time of how the candidates were performing according to users on Twitter.

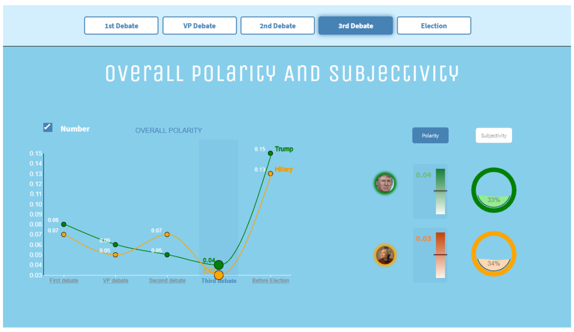


Figure 2: Time Series

5.5.3 *Bar Chart*

A bar chart was created to demonstrate the key election topics at various points through the election. It demonstrates that the key topic that seemed to arise in the second debate was "Women". Over the five key points and fifteen available slots it was "Make America Great Again" that appeared on four of five possible occasions, "Jobs" appeared three times, "Racial Issues" also appearing three times, "Women" and "Bhengazi" twice and "Tax" just the one solitary occasion.

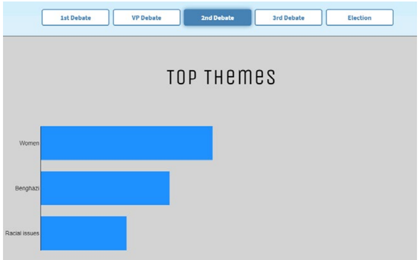


Figure 3: Bar Chart of Top Themes

5.5.4 *Themes Grid*

A top themes grid was developed in order to visualize the top trending themes at the five key stages of the election. This grid allows the user to observe the top three themes by State at each of the five milestones.

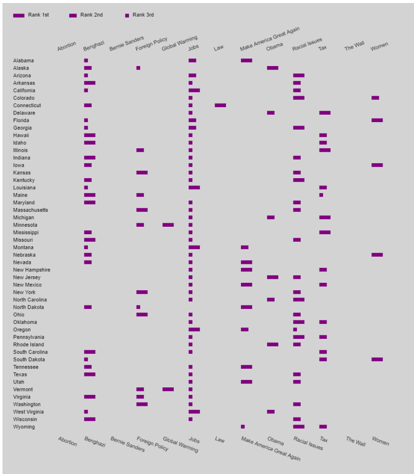


Figure 4: Top Themes Grid

5.5.5 *Force Directed Graph*

One of the key areas that is beneficial to explore in the interconnected world of Social Media is how users on the platform are connected. To do this a force directed graph was created. What this demonstrated is that the network and mentions of the Twitter account realDonaldTrump, where there were over 30 tweets by a user, grew substantially over the course of the election which was initially smaller in size than that of the account for HilaryClinton. Noting that from the outset realDonaldTrump seemed to have a larger network, particularly where there were a small number of tweets.

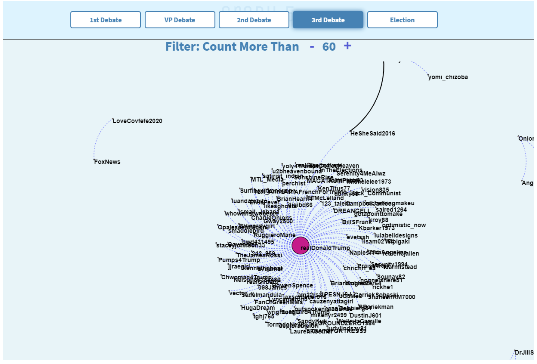


Figure 5: Force Directed Graph

5.5.6 *Emoji Analysis*

Whilst the emoji analysis had been built into the sentiment analysis, it was identified that a valuable exercise would be to undertake a review of the emoji's used. The emoji's had previously been tagged into five key categories being: disgust, sadness, uncertainty, joy, and surprise. There is movement in the use of certain emoji's and categories of the emoji's based on the outcomes at the five key points in the time series that can be seen using three graphs on the Tweetstorm platform including the spiderweb chart, a pie chart and a dendrogram. All these charts are displayed for the user who is then able to move through the various points to see any change.

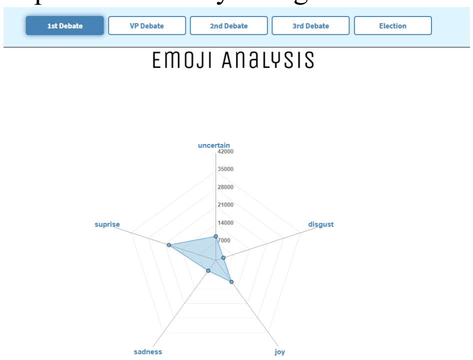


Figure 6: Emoji Spider Web Analysis

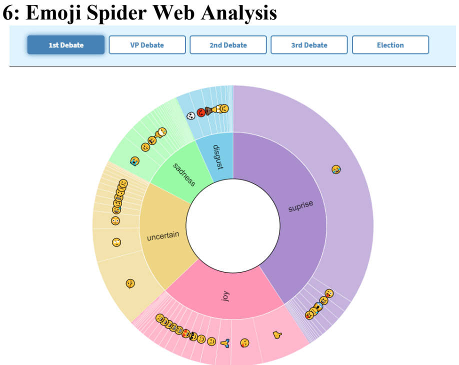


Figure 7: Emoji Use Pie Chart

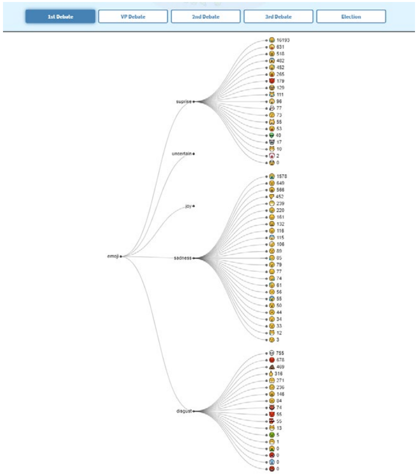


Figure 8: Emoji Dendrogram

5.6 Presentation

A web application was developed using HTML, CSS and JavaScript which is hosted on a server. The simple technology diagram is outlined in figure 6.

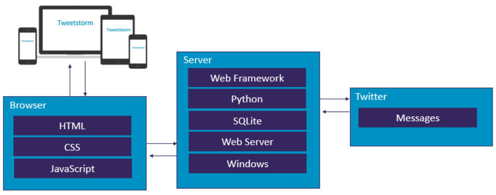


Figure 9: Technology Stack

The web page that is displayed will outline the purpose of the platform and how it is to be used before presenting a series of filters and options to explore the data visualizations.

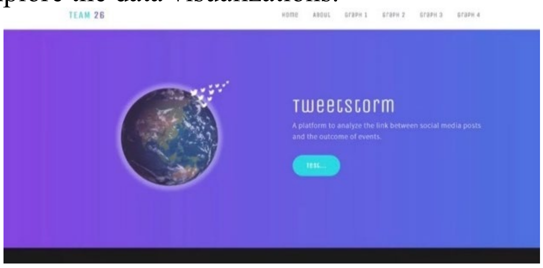


Figure 10: Tweetstorm home page

5.7 Intuition

There are several reasons that this approach is leading others. The reasons include:

- The cleansing approach taken to ensure the success of the sentiment modeling is extremely comprehensive, structuring to provide a clear, clean dataset for sentiment modeling;
- The sentiment and polarity scoring expand beyond previous models providing a wider scoring model to gain a deeper understanding of the leaning of the tweets;
- the processing and application of sentiment on a combination of characters to determine a score for the use of emoji's; and
- the approach to visualization for the easy interpretation and exploration of the data for all potential audiences.

6. Experiments and Observations

In order to develop our platform, we conducted several experiments. The goal of these experiments was to ensure that sentiment scores were meaningful, the visualizations detailed the appropriate story, and the topics extracted from the data were of value. The experiments that we have performed have been undertaken through querying and cleaning our tweet data on SQLite as well as using Python in making the API calls to the Twitter API and conducting cleaning activities in the processing of data prior to its storage. The details are explained below.

6.1 Experiments

We conducted a series of experiments that were designed to answer the following questions. Can we get a satisfactory performing sentiment analysis model on a series of tweets? Are we able to satisfactorily classify and obtain sentiment from emoji's? Are we able to visualize in a user-friendly and easy to understand manner, the outcomes of our model? Can we prepare the tweet data to an appropriate level for our models to work well?

The experiments conducted in the development of the model are detailed as follows.

6.1.1 Sentiment and Polarity Models

The key to the Tweetstorm platform is the production of appropriate sentiment, polarity and impact measurements of each tweet. The scoring includes the training of the model on a series of pre-established words and phrases. Key improvements to the algorithm and model were dependent on the text preprocessing, thus the sentiments distributions were analyzed before and after each change. Furthermore, we looked at the word cloud of positive and negative tweets from our model to evaluate whether the keys words aligned with the positive and negative polarities expected.

In initial experiments on the raw data it was identified that there were a large proportion of tweets that had 0 polarity and subjectivity. The word cloud also had no meaningful words. In attempting to address this issue the team focused on cleansing the data. The iterations of the experiment and outcomes are detailed as follows:

#	Experiment	Outcome
1	Remove stop words, remove tags and links, remove non-	Minimal improvement. Word cloud have non-

	alpha-numeric characters.	meaningful characters.
2	Remove retweet labels, convert tweets to lowercase. Lemmatize all tweets using POS tagging.	Major improvement to sentiment. Word cloud is significantly more understandable now, with most words being useful.
3	Remove the word ‘debate’.	Minimal improvement
4	Combine words that mean the same thing.	Major improvement to sentiment. Donald Trump and Hillary Clinton are now mostly grouped together and not double counted.
5	Remove https as these are still present after tag removal.	Minimal improvement

6.1.2 Classifying Emoji’s

How do you obtain sentiment for a combination of characters rather than text? This was a key question for the team, after a series of experiments this piece has been refined to allow accurate processing of the data into the scoring model built. The experiments included:

#	Experiment	Outcome
1	Sentiment modelling based on text in tweets.	Problematic due to a number of factors including emoji only tweets.
2	Emoji emotion tags.	Tagging an emotion to an emoji type vastly improved the performance.
3	Sarcasm identification	These tests were not completed to a satisfactory level.

Visualization Testing

There were several experiments that were required to be conducted on ensuring that the visualizations were function for the user, these included:

6.1.2.1 Word Cloud

#	Experiment	Outcome
1	disable colocations in word cloud to remove duplicate words	Word cloud does not have repeated words

The word cloud was not used in the final version of the platform however the word cloud experiments assisted with ensuring the top themes analysis was ready without need for any further clean-up.

6.1.2.2 Geographic Polarity

#	Experiment	Outcome
1	Develop dynamic movement based on page filters	Smoother transition of the visualizations

There were several lessons learnt through the development of this visualization, particularly in the of the page and map filters which generated the details in other sections of the page.

7. Team Member Contributions

Throughout the development of this platform all team members have contributed a similar amount of effort in the design and build.

The original structure of work was as follows:

Role	Tasks	Team Member/s	Time Allocation
Coordinator	• Ensuring team deliverables are met • Coordination of check-ins • Report drafting and submission	Matthew Tinker	0.5
Data Acquisition	• Identifying suitable datasets • Preparing and curation of data • Formatting of data	Jeffrey Lee Minhao Leong	0.5 each
Model Build	• Development of analytic models	Jeffrey Lee Minhao Leong	0.5 each
Visualization	• Development of interactive visualizations • Query backend datasets	Jinchao Lin Jun Xiong Tan Matthew Tinker	Up to 1.0 each
Web Interface	• Development of web interface	Jun Xiong Tan Matthew Tinker	0.25 each

The work structure and progress remained in line with the original plan with all key milestones were met.

8. Discussion

The Tweetstorm platform provides its users with an easy to use interface that allows them to see how influential messages on the platform can be on event outcomes. This is what this project set out to do and it has achieved.

Whilst the model did not correctly predict the eventual winner in every state, it did clearly demonstrate through tweet sentiment analysis that the likely candidate to be successful in their bid for the US Presidency was Donald Trump and only at one stage did the sentiment analysis suggest that he was not the preferred candidate.

This first iteration focused on the Twitter activity related to the 2016 US Presidential election however the model development and the user interface are adaptable for other use cases in further iterations.

9. Conclusion

This document has provided an outline of the Tweetstorm platform. Team 26 set out to develop a platform that, after conducting sentiment analysis on tweets, would allow its users to see the influence of social media on the outcome of events. In developing this innovative solution to the problem identified, Team 26 has met its initial objectives and propose that Tweetstorm is a platform that could be adapted for many future use cases so that people can obtain this understanding of the influence of the consumption of content from social media platforms.

10. References

[1] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-36.

[2] Chen, Y., Yuan, J., You, Q., & Luo, J. (2018, October). Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In Proceedings of the 26th ACM international conference on Multimedia (pp. 117-125).

[3] Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), 1-41.

[4] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.

[5] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. Mathematical and Computational Applications, 23(1), 11.

[6] Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. Social Media+ Society, 3(1), 2056305117691545.

[7] Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access, 5, 2870-2879.

[8] Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im "predicting elections with twitter: What 140 characters reveal about political sentiment". Social science computer review, 30(2), 229-234.

[9] Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. European Journal of Communication, 32(3), 257-273.

[10] Khalid, A. (2013). Systems Engineering Graduate Research as Part of Curriculum–Summary of Research. Procedia Computer Science, 16, 967-975.

[11] Lerman, K., & Ghosh, R. (2010, May). Information contagion: An empirical study of the spread of news on digg and twitter social networks. In Fourth International AAAI Conference on Weblogs and Social Media.

[12] Littman, J., Wrubel, L., Kerchner, D. (2016). 2016 United States Presidential Election Tweet ID's, https://doi.org/10.7910/DVN/PDI7IN, Harvard Dataverse, V3, accessed 25 March 2020.

[13] Liu, K. L., Li, W. J., & Guo, M. (2012, July). Emoticon smoothed language models for twitter sentiment analysis. In Twenty-sixth aAAI conference on artificial intelligence.

[14] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010, May). From tweets to polls: Linking text sentiment to public opinion time series. In Fourth international AAAI conference on weblogs and social media.

[15] Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. Information Processing & Management, 53(4), 764-779.

[16] Rakhmetullina, A., Trautmann, D., & Groh, G. (2018). Distant Supervision for Emotion Classification Task using emoji2emotion. In Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018). Stanford, CA, USA. http://ceurws. org (Vol. 2130).

[17] Sari, Y. A., Ratnasari, E. K., Mutrofin, S., & Arifin, A. Z. (2014). User emotion identification in twitter using specific features: Hashtag, emoji, emoticon, and adjective term. Jurnal Ilmu Komputer dan Informasi, 7(1), 18-23.

[18] Singh P., Sawhney R.S. (2018) Influence of Twitter on Prediction of Election Results. In: Saeed K., Chaki N., Pati B., Bakshi S., Mohapatra D. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol 564. Springer, Singapore.

[19] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010, May). Predicting elections with twitter: What 140 characters reveal about political sentiment. In Fourth international AAAI conference on weblogs and social media.

[20] Van Hee, C., Lefever, E., & Hoste, V. (2018, June). Semeval-2018 task 3: Irony detection in English tweets. In Proceedings of The 12th International Workshop on Semantic Evaluation (pp. 39-50).

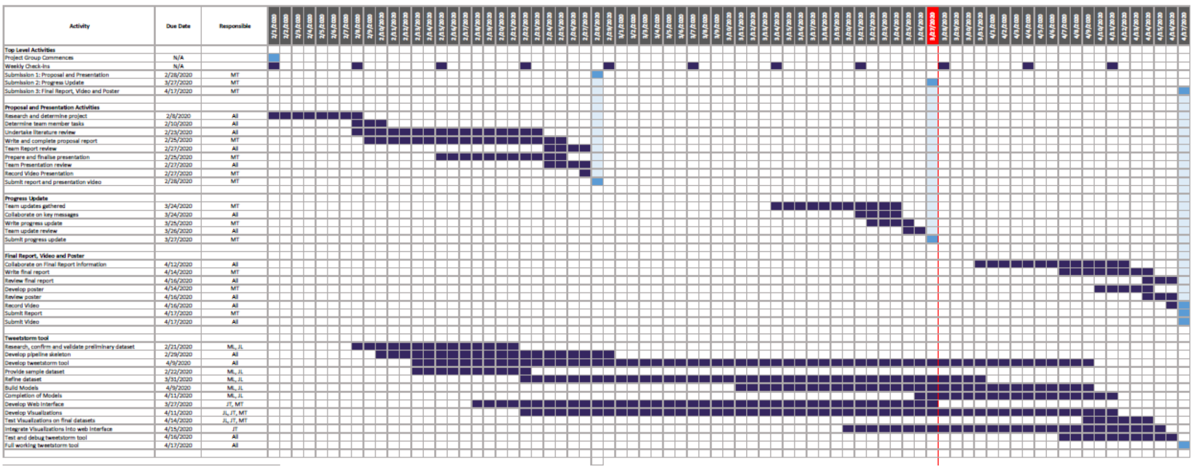
[21] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 system demonstrations (pp. 115-120). Association for Computational Linguistics.

[22] Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social media as information

source: Recency of updates and credibility of information. Journal of computer-mediated communication, 19(2), 171-183.
[23] Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 35(3), 399-433.
[24] Wolny, W. (2016). Emotion analysis of twitter data that use emoticons and emoji ideograms.
[25] Wong, F. M. F., Tan, C. W., Sen, S., & Chiang, M. (2013, June). Quantifying political leaning from tweets and retweets. In Seventh International AAAI Conference on Weblogs and Social Media.
[26] Wood, I., & Ruder, S. (2016, May). Emoji as emotion tags for tweets. In Proceedings of the Emotion and Sentiment Analysis Workshop LREC2016, Portorož, Slovenia (pp. 76-79).
[27] Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019, May). Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 218-226).
[28] Zhao, W. X., Jiang, J., He, J., Song, Y., Achananuparp, P., Lim, E. P., & Li, X. (2011, June). Topical keyphrase extraction from twitter. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1 (pp. 379-388). Association for Computational Linguistics.

11. Appendix

11.1 Project Schedule



11.2 Detailed Activities

Activity	Due Date	Responsible
Top Level Activities		
Project Group Commences	Complete	
Weekly Check-ins	Complete	
Submission 1: Proposal and Presentation	Complete	MT
Submission 2: Progress Update	Complete	MT
Submission 3: Final Report, Video and Poster	Complete	MT
Proposal and Presentation Activities		

Activity	Due Date	Responsible
Research and determine project	Complete	All
Determine team member tasks	Complete	All
Undertake literature review	Complete	All
Write and complete proposal report	Complete	MT
Team Report review	Complete	All
Prepare and finalize presentation	Complete	MT
Team Presentation review	Complete	All
Record Video Presentation	Complete	MT
Submit report and presentation video	Complete	MT
Progress Update		
Team updates gathered	Complete	MT
Collaborate on key messages	Complete	All
Write progress update	Complete	MT
Team update review	Complete	All
Submit progress update	Complete	MT
Final Report, Video and Poster		
Collaborate on Final Report information	Complete	All
Write final report	Complete	MT
Review final report	Complete	All
Develop poster	Complete	MT
Review poster	Complete	All
Record Video	Complete	All
Submit Report	Complete	MT
Submit Video	Complete	All
Tweetstorm tool		
Research, confirm and validate preliminary dataset	Complete	ML, JL
Develop pipeline skeleton	Complete	All
Develop tweetstorm tool	Complete	All
Provide sample dataset	Complete	ML, JL
Refine dataset	Complete	ML, JL
Build Models	Complete	ML, JL
Completion of Models	Complete	ML, JL
Develop Web Interface	Complete	JT, MT
Develop Visualizations	Complete	JL, JT, MT
Test Visualizations on final datasets	Complete	JL, JT, MT
Integrate Visualizations into web interface	Complete	JT
Test and debug tweetstorm tool	Complete	All
Full working tweetstorm tool	Complete	All

11.3 Work Breakdown

Role	Tasks	Team Member/s	Time Allocation
Coordinator	• Ensuring team deliverables are met • Coordination of check-ins • Report drafting and submission	Matthew Tinker	0.5
Data Acquisition	• Identifying suitable datasets • Preparing and curation of data • Formatting of data	Jeffrey Lee Minhao Leong	0.5 each
Model Build	• Development of analytic models	Jeffrey Lee Minhao Leong	0.5 each
Visualization	• Development of interactive visualizations • Query backend datasets	Jinchao Lin Jun Xiong Tan Matthew Tinker	Up to 1.0 each
Web Interface	• Development of web interface	Jun Xiong Tan Matthew Tinker	0.25 each

11.4 Technology Diagram