# A sentiment analysis of the Black Lives Matter movement using Twitter

10.17975/sfj-2022-015

Jacqueline Peng[1], Jun Shen Fung[2], Muhammad Murtaza[3], Afnan Rahman[4], Pallav Walia[5], David Obande[6], and Anish R. Verma[7]

[1] Biomedical Graduate Studies, University of Pennsylvania, Pennsylvania, United States of America

[2] Department of Biomedical Engineering, University of British Columbia, British Columbia, Canada

[3] John Fraser Secondary School, Ontario, Canada

[4] Department of Physical and Mathematical Sciences, University of Toronto, Ontario, Canada

[5] Department of Management Sciences, University of Waterloo, Ontario, Canada

[6] Department of Population Medicine, University of Guelph, Ontario, Canada

[7] 1QB Information Technologies (1QBit), British Columbia, Canada

**ABSTRACT:** As more attention is brought to the issue of racial injustice, public sentiments and opinions on racial issues are increasingly important to track. At the same time, recent progress in machine learning and natural language processing methods, coupled with the growing amount of available data for training and analysis, allows researchers to extract sentiments from text data at large scales. We applied a natural language processing framework to study public sentiment surrounding the Black Lives Matter (BLM) movement. Specifically, we used a state-of-the-art BERT model fine-tuned for Twitter sentiment classification to predict the sentiment from approximately 1 million tweets from July 2013 to March 2021 related to BLM. The BERT model was trained on the Sentiment 140 dataset on which it obtained an AUC of 0.97 on the training data and 0.94 on testing data, outperforming other machine learning models. We found that retweet frequency and word count frequency were able to illustrate important themes in the BLM movement as well as indicate events of significant importance to the movement. Additionally, sentiment analysis revealed which of these themes and events were associated with positive public sentiment, such as social justice, and which were associated with negative sentiment, such as police brutality. Our analyses can also be applied to better understand other social and political movements to aid related research and activism.

**KEYWORDS:** sentiment analysis, Black Lives Matter, machine learning, Twitter

## INTRODUCTION

The Black Lives Matter (BLM) movement is a decentralized grassroots movement that aims to spotlight the systemic and institutionalized oppression and racially instigated violence on the black community [1]. The BLM movement is affiliated with the Black Lives Matter Global Network Foundation whose purpose is to diminish white supremacy and build local power to voice the rights and social justice for the historically marginalized black community [1]. While the BLM movement has been around since 2013, it has gained a substantial amount of attention and momentum in recent years, especially since May 25, 2020, after the brutal murder of George Floyd by then Minneapolis police officer Derek Chauvin. Social media has served to propel the BLM movement, as well as other similar movements seeking racial justice for those in the black community, through hashtags like #BlackLivesMatter, which as of 2018 has been tweeted 30 million times [2].

This study aims to understand (a) what themes are associated with the BLM movement and sentiments around them, and (b) how sentiment of the BLM movement has changed over time and with important events. A better understanding of the BLM movement can help the academic community engage in research to promote racial justice and, more broadly, social justice issues.

Due to the influence of social media on the BLM, there is a wealth of information and insights that can be drawn

STEM Fellowship Journal



by analyzing posts on social media platforms like Twitter, Facebook, and Instagram. Twitter, specifically, is an exceptionally useful source of data for analyzing social and political movements because of its retweet and hashtag features that make it easy to track down the popularity and spread of a movement, cause, or idea. Automated methods are required in the text-based analysis of social media data because it would be too costly and time-consuming for humans to manually read and annotate such a vast amount of social media texts. In recent years, the development of natural language processing (NLP) methods has made it possible to utilize computational power to automate textual analysis [3]. One such NLP method is sentiment analysis, which can be used to determine whether the sentiment of a text is positive or negative. Sentiment analysis is useful in tracking public perceptions of certain events, people, and ideas. Previous studies have developed well-established pipelines for performing sentiment analysis on Twitter data, and some examples which use cases include analyzing public sentiment towards the 2012 United States (US) election candidates [4] and, recently, the public perception of the COVID-19 pandemic on Twitter [5]. There has been a previously published study that applied sentiment analysis to study BLM, however, it did not use a state-of-the-art transformer model as we have done, the size of the dataset is 10 times smaller, and it did not examine sentiment over time [6]. Other literature in this area focus primarily on the generation of BLM data rather than drawing conclusions [7, 8].

Thus, we use NLP and sentiment analysis to better understand the sentiments around, themes, and propagation of the BLM movement, like analyses in the literature for other topics. The analyses demonstrated in this paper can also be applied to better understand other social and political movements.

# METHODS

## Data collection

The Twitter API (version 2) was used to collect tweets related to BLM. This was done using the query '(@Blklivesmatter OR #blacklivesmatter) lang:en -is:retweet', which searches for tweets mentioning the official BLM Twitter account (@Blklivesmatter) and tweets with the BLM hashtag (#blacklivesmatter), and only returns tweets in English. Retweets are filtered out of these data.

We applied this search query daily from July 2013, when the BLM Twitter account was created, up to and excluding March 1, 2021. The Twitter API only returned a maximum of 500 tweets each day due to constraints on the API. We recognize this tweet cap is a limitation of our analysis. This provided a large enough sample to approximate major trends in the complete BLM tweet data. However, we acknowledge future analysis should be conducted with a larger corpus. In total 996,624 unique tweets related to BLM were collected, with the earliest tweet occurring July 14, 2013, and the most recent tweet occurring February 28, 2021. We will refer to this dataset as the BLM Twitter corpus.

## Sentiment analysis model training

The Sentiment140 dataset [9] was used to train five machine learning models for a sentiment analysis task: logistic regression, naive Bayes, XGBoost [10], LSTM [11], and BERT [12]. The sentiment analysis task was a supervised binary classification problem where either a label of positive or negative sentiment is predicted for each tweet. The Sentiment140 dataset has 1.6 million tweets which were used to train the models. The scikit-learn Python package (https://scikit-learn.org/) [21] was used to train the logistic regression and naive Bayes classifiers, the XGBoost Python package (https://xgboost.readthedocs.io/) [10] was used to train the XGBoost classifier, and Keras (https://keras.io/) [22] was used to train the LSTM classifier. A pre-trained version of BERT [12] was obtained using the 'bert-base-uncased model' from the HuggingFace transformers Python package (https://huggingface.co/transformers/) [23] and fine-tuned on the Sentiment140 dataset using Tensorflow (https://www.tensorflow.org/) [24]. The Sentiment140 dataset was split into two sets of 90% and 10% for training and testing, respectively. Where applicable, hyperparameter tuning was performed using 5-fold cross-validation for all models, except for the deep learning models (LSTM and BERT) for which one cross-validation fold was used, due to the time-consuming nature of deep learning model training. The models were evaluated using the AUC-ROC metric. AUC-ROC is commonly used to compare how accurately the ML models predict the sentiment of new tweets, considering both sensitivity and specificity, and represents the area under the ROC curve, which plots the rate of true positives against the rate of false positives. Sentiment analysis model training details are provided in the Appendix and a graphical overview of the methods is shown in Figure 1.

## RESULTS

The performance of the five models on the training and testing datasets are shown as ROC curves in Figure 2 A) and B), respectively. The performance of the models is also quantified in Table 1. All models perform relatively well on the training and testing sets; all area AUC values are greater than 0.86 and some are greater than 0.9. As expected, the models perform equally as well or better on the training data compared to the testing set. While the three machine learning models, logistic regression, naive Bayes, and XGBoost, perform slightly differently on the training set, they perform very similarly on the testing set. The deep learning models, LSTM and BERT, perform better on both the training and testing sets than the three machine learning models, with the BERT model performing the best out of all five models. As the best performing model, BERT was used to label the sentiment of the curated BLM Twitter corpus, consisting of approximately 1 million tweets from July 2013 to March 2021 related to BLM.

To extract common themes from the BLM Twitter corpus, we generated a word cloud to show the 500 most common words in the corpus (Figure 3). The most frequent word in the corpus is ‘blacklivesmatter’ with 945,895 instances out of the nearly 1 million unique tweets. The next most frequent words are ‘black’, ‘people’, ‘police’, and ‘white’. Additionally, Figure 4 visualizes the top 25 most common words in the BLM Twitter corpus. The 25 words are categorized into common themes, such as advocacy, government/politics, and racial injustice.

To explore sentiment over the various themes in the BLM Twitter corpus, we compared words found primarily

**Table 1.** The performance of five machine learning/deep learning models on the Sentiment140 dataset are given below. The metrics reported include precision, recall, F1 score, and AUC-ROC, all which are common metrics used for binary sentiment classification tasks. The first number in each cell represents the performance of the model in the training set and the second number represents the performance of the model on the testing set.

| Model | Precision | Recall | F1 score | AUC-ROC |
|---|---|---|---|---|
| Logistic regression | 0.83/0.80 | 0.80/0.78 | 0.82/0.79 | 0.89/0.87 |
| Naive Bayes | 0.80/0.73 | 0.88/0.81 | 0.84/0.77 | 0.92/0.86 |
| XGBoost | 0.67/0.68 | 0.84/0.84 | 0.75/0.75 | 0.87/0.87 |
| LSTM | 0.88/0.83 | 0.89/0.84 | 0.88/0.83 | 0.95/0.91 |
| BERT | 0.92/0.88 | 0.91/0.86 | 0.91/0.87 | 0.97/0.94 |

Precision represents the number of true positives out of all the tweets predicted positive by the model, and the equation for it is (# of true positives)/ (# of true positives + # of false positives). Recall represents the number of true positives predicted by the model out of all the tweets that are actually positive, and the equation for it is (# of true positives/ (# true positives + # false negatives). F1 score represents the balance between precision and recall and the equation for it is (2 x precision x recall)/ (precision + recall).

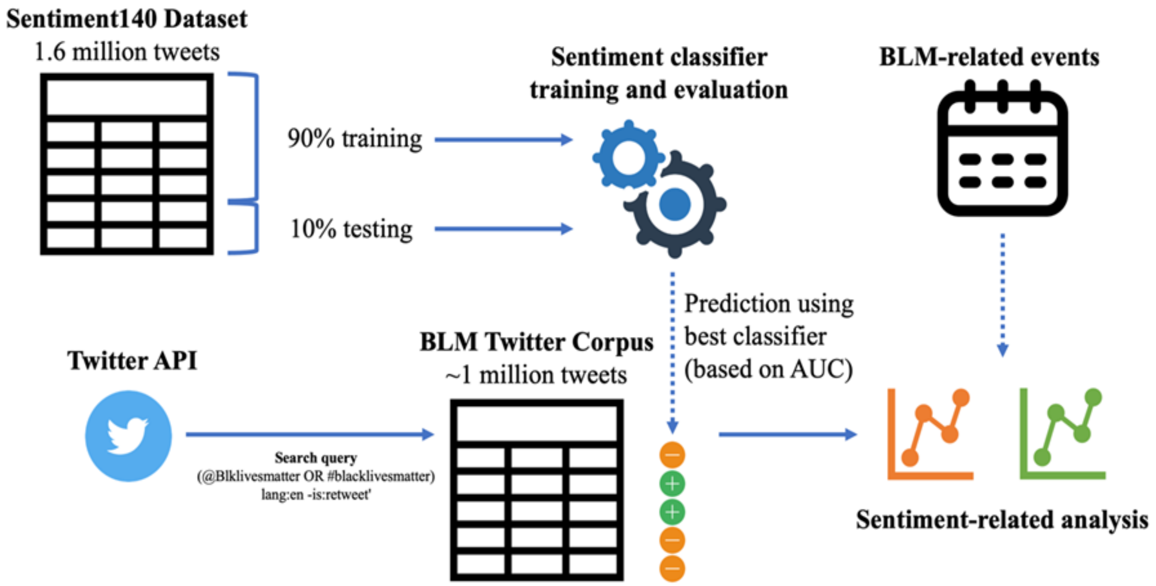


**Figure 1.** Methods overview. The goal of sentiment analysis in this study was to accurately automate the classification of tweets as positive or negative sentiment. To do this, several commonly used machine learning (ML) models were trained on 90% of tweets annotated as either positive or negative sentiment in the Sentiment140 dataset [9]. The best model was chosen based on the AUC-ROC metric on the remaining 10% of the dataset. The best model was then applied on ~1 million BLM-related tweets (i.e., the BLM Twitter corpus), collected using the Twitter API, to annotate the sentiment of these tweets. Downstream analysis was performed using these annotations, including correlating sentiment with BLM-related events.

STEM Fellowship Journal



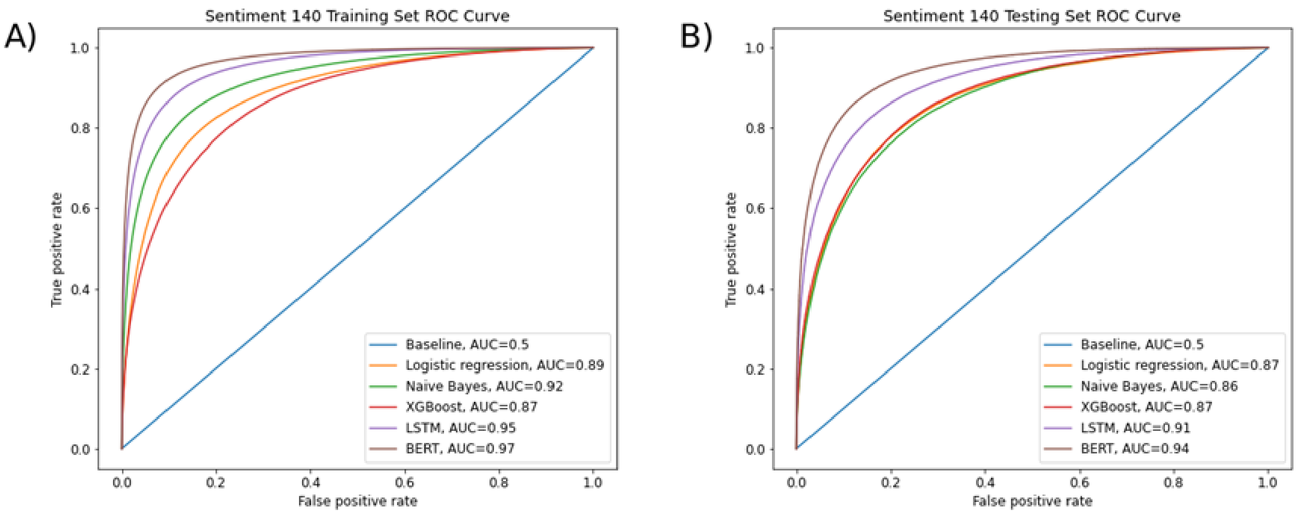


**Figure 2.** Performance of five machine learning/deep learning models on the Sentiment140 dataset. Five machine learning models- logistic regression, naive Bayes, XGBoost, LSTM, and BERT- were trained on 90% of the Sentiment140 dataset and tested on the remaining 10% of the dataset. The task the models were trained to perform was binary classification of the tweets in the dataset based on their sentiment. The dataset has a total of 1.6 million tweets and corresponding binary labels, representing positive or negative sentiment. Each curve is an ROC curve for a single model. The baseline model predicts the majority class. The AUC value is displayed in the legend for each model. A) shows the results on the training dataset and B) shows the results for the testing dataset.

in positive sentiment tweets and words found primarily negative sentiment tweets, using labels from the trained BERT model. In Figure 5, two independent word clouds were generated to visualize the top 200 most frequent words from tweets predicted as having positive or negative sentiment.

We also wanted to investigate how various social, political, and other events affected the retweet frequency of tweets in the BLM Twitter corpus since the BLM Twitter account (@ Blklivesmatter) was established. Relevant events were found by conducting textual analysis on the aggregated tweets and by matching tweet content with key incidents that occurred during the associated week. Retweet frequency over time is shown in Figure 6, annotated with relevant events.

To track sentiment over time and its relationships with the various BLM-related events, a similar version of Figure 6 was plotted but with the positive and negative tweets separated out, to create Figure 7. Figure 7 demonstrates that retweet spikes related to events involving shootings or deaths often stem from tweets predicted to have negative sentiment.

## DISCUSSION

Out of the five trained models, the three machine learning models (i.e., logistic regression, naive Bayes, XGBoost) performed similarly to each other on the testing set, while the deep learning models (i.e., LSTM, BERT) performed better on both the training and testing sets. This is consistent with previous studies that have shown that, given enough data, deep learning models are generally better than traditional machine learning models for NLP tasks, including sentiment analysis, because of their ability to encode the sequential nature of language [13]. It is not surprising that the BERT model performed the best, because of the ability of the state-of-the-art transformer model to capture contextual cues in texts using the attention mechanism [14]. Further, BERT is a pre-trained model, having already been trained on a huge amount of text data beyond Sentiment140. Indeed, the Sentiment140 dataset is only used to fine-tune the model. Therefore, BERT can learn and generalize better than the other models for our Twitter sentiment analysis task.

We used word clouds to visualize important words and themes related to the BLM movement. The most common words and the word themes in Figure 3 suggest that recent events, including the George Floyd murder, had a significant impact on the topics that the public were tweeting as well as the overall public sentiment regarding BLM. The most common word 'blacklivesmatter' makes sense as the top word because it comes from #blacklivesmatter, which was used to retrieve the corpus' tweets.

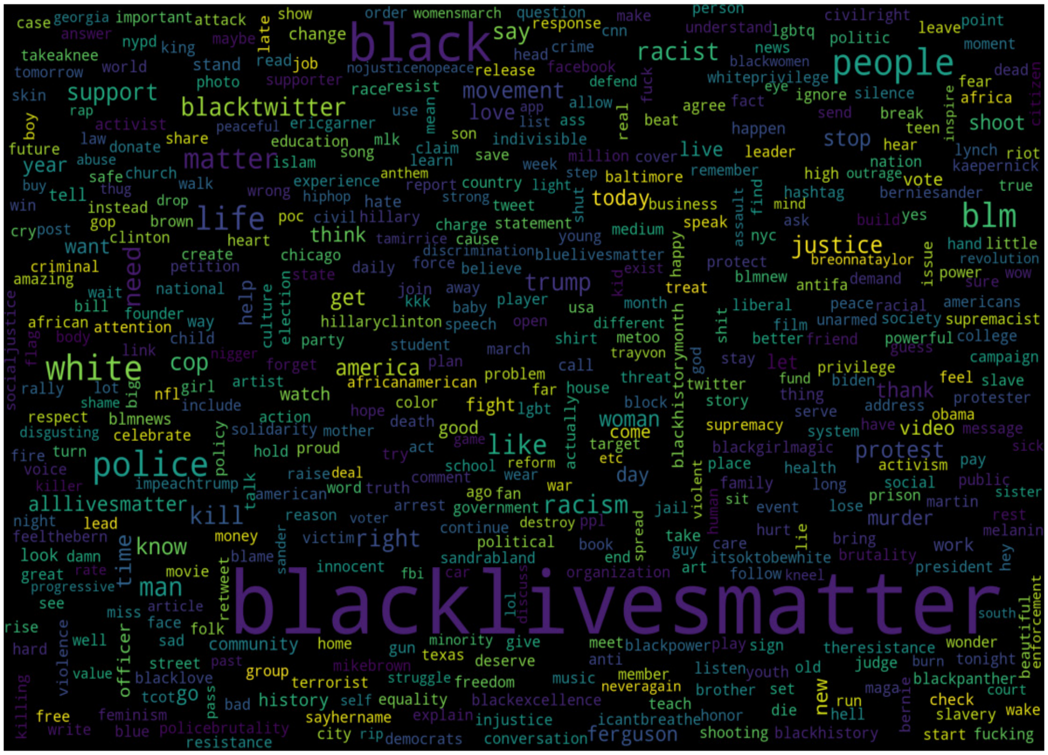

**Figure 3.** Word cloud of BLM Twitter corpus. After pre-processing and tokenizing the BLM Twitter corpus, scikit-learn's CountVectorizer was used to count the frequency of each token/word in the corpus. A word cloud was generated using the wordcloud Python package (https://amueller.github.io/word_cloud/). This figure shows the 500 most common words in the BLM Twitter corpus. The size of a word in the word cloud is scaled to its frequency in the corpus.

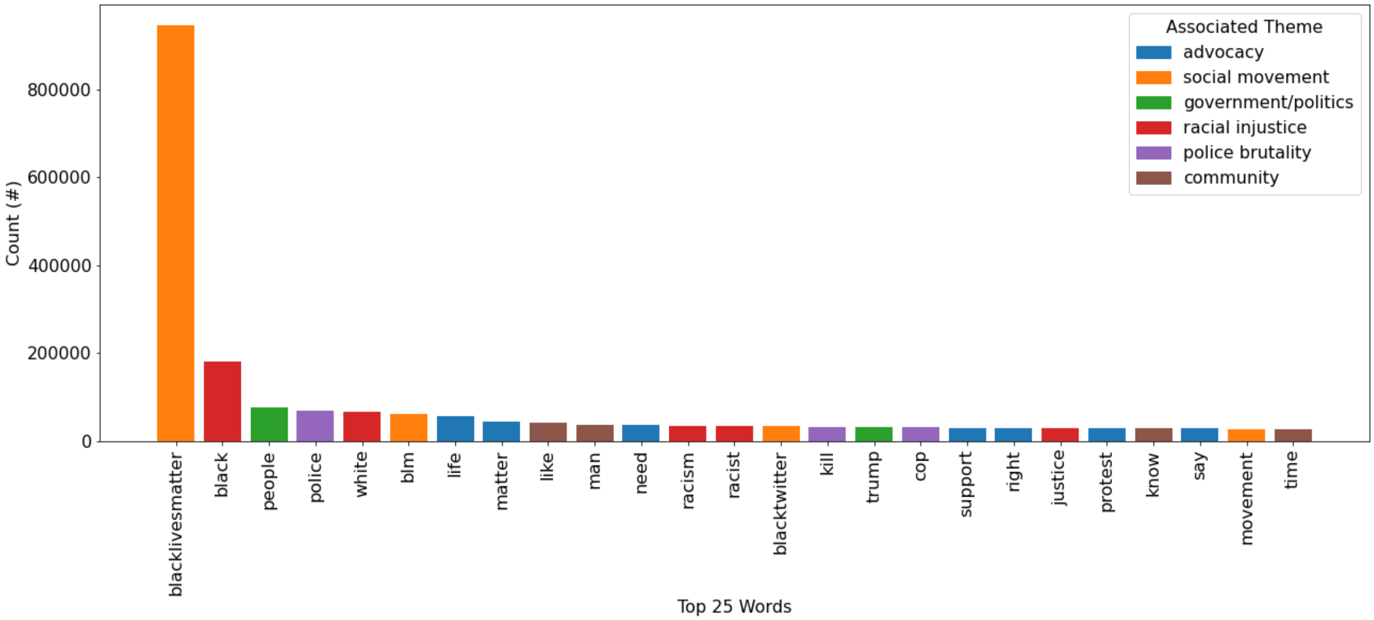


**Figure 4.** Top 25 most frequent words in the BLM Twitter corpus. The bar graph plots the top 25 words in the BLM Twitter corpus with respect to their frequency (count of how many times they appear in the corpus). Each word was manually assigned an associated theme and each theme is given a different color in the bar graph.

STEM Fellowship Journal



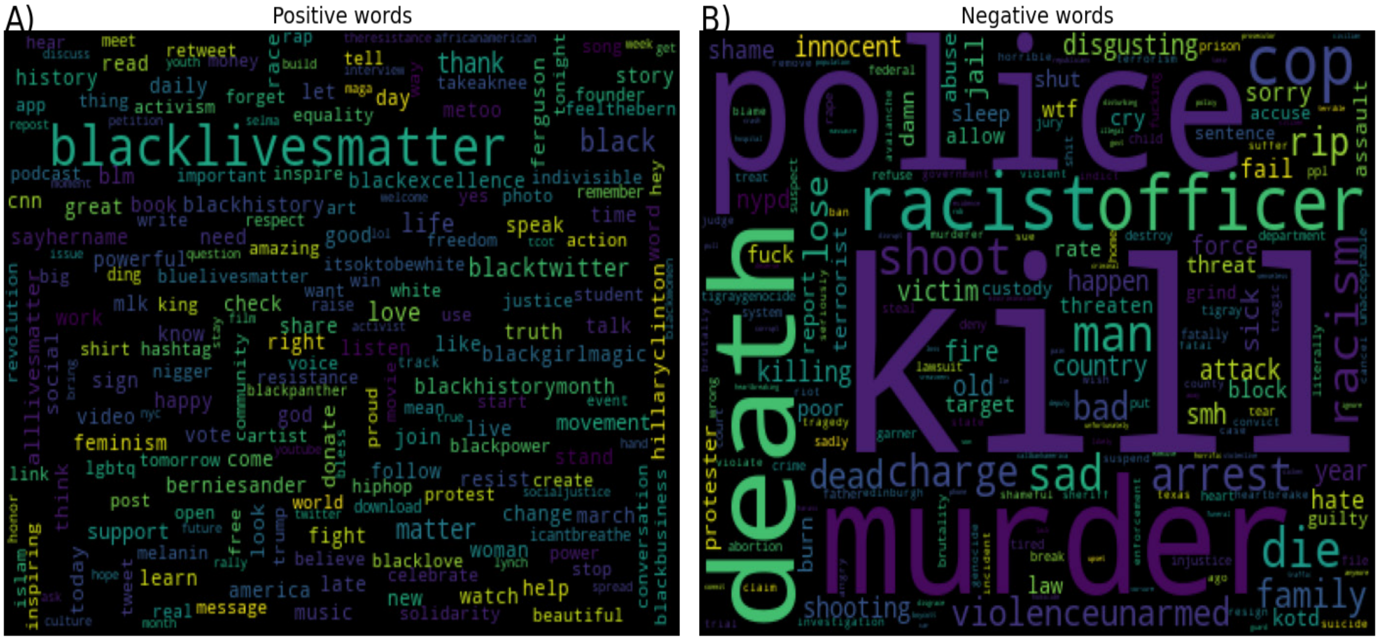


**Figure 5.** Word cloud of BLM Twitter corpus for A) positive sentiment tweets and B) negative sentiment tweets. Tweets in the BLM Twitter corpus predicted as having positive sentiment with a confidence level greater than 0.8 by the trained BERT sentiment analysis model were deemed “positive tweets” and tweets predicted as having negative sentiment with a confidence level greater than 0.8 were deemed “negative tweets”. To take the overall frequency of a word into account and discriminate between how positive and negative it is in terms of sentiment in the BLM Twitter corpus, the word cloud in A) scales each word by its frequency among the positive tweets minus its frequency among the negative tweets, and vice versa for the word cloud in B). Therefore, the interpretation is that the size of a word in A) scales with how positive it is, while the size of a word in B) scales with how negative it is.

The theme of police brutality is suggested by words such as ’police’, ‘cop’ and ‘kill’ in the word cloud. Certain politicians are also frequent in the word cloud, including former President Donald Trump (‘trump’). President Trump was relevant to the BLM movement because of his opposition against BLM protests and failure to support the movement in general [15]. Finally, words such as ‘black’, ‘white’, ‘racism’, ‘racist’, ‘support’, ‘justice’, ‘protest’, perfectly align with the BLM movement goals, and these words show how people are using Twitter as a platform to advocate for the Black community. While word clouds are good for modelling the importance of words independently, they do not take contextual information into account (i.e., which words are used together). A future research direction from this study is to use topic modelling [16] to automatically detect themes from the tweets, and which words belong in these themes, based on how certain words are used together or in similar contexts.

The analysis of words, separated by their association with positive sentiment or negative sentiment labeled tweets, as shown in Figure 5, reveals a positive sentiment for the BLM movement and a negative reaction towards police, killing, and racism. Specifically, through analyzing and classifying the sentiment of public opinions in the form of tweets on Twitter, negative words (e.g., ‘police’, ‘racist’, ‘officer’, ‘kill’) are shown to align with public news about police brutality, murdering, and unfair treatment towards black minorities. This negative sentiment towards the policing and treatment of black individuals is consistent with recent research [17]. On the other hand, democratic politicians, such as Bernie Sanders (‘berniesanders’) and Hillary Clinton (‘hillaryclinton’), appear in the positive word cloud, which aligns with their activism against racism and their support of progressive ideas and movements, including BLM [18]. Twitter as a social media platform allows public opinion to become apparent and for the surveying and sampling of tweets encompassing positive encouragement of the BLM movement, as well as negative sentiment towards racism, social injustices, and criticism of law enforcement.

The BLM movement has largely utilized social media platforms like Twitter to spread news and to bridge the gap in mainstream narratives. There is a clear development in social media participation to protest individuals, groups, and organizations, and incite legal action [19]. Most of the tweets were found to address racial bias in the US law

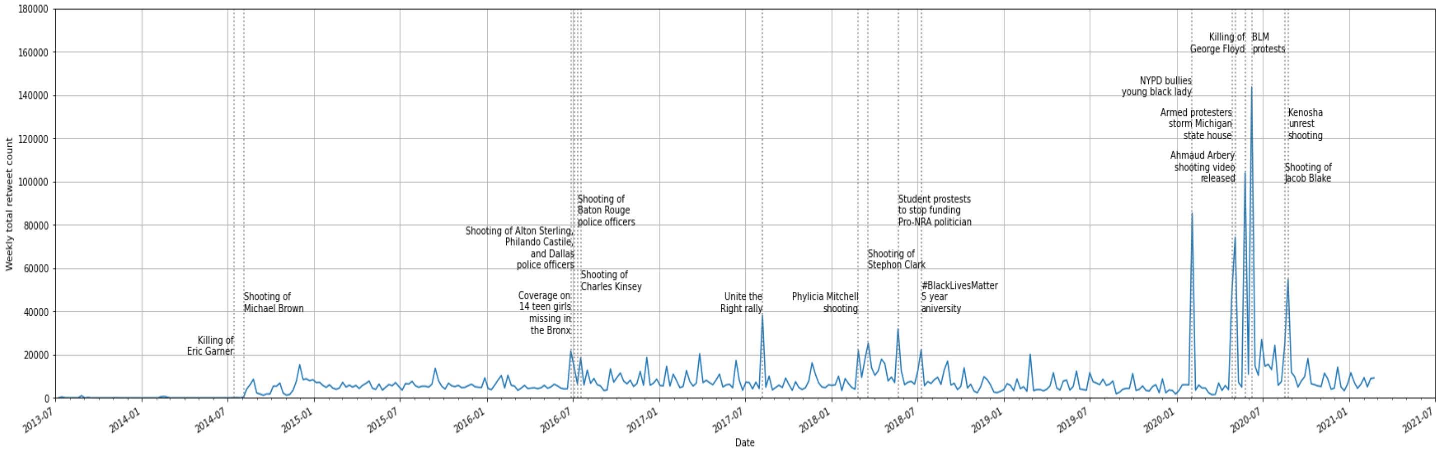


**Figure 6.** BLM Twitter corpus retweets over time with annotations. The figure shows the aggregated weekly total retweet count from July 2013 to March 2021 of the tweets in the BLM Twitter corpus. That is, the tweets were grouped by the start date of the week the tweets were created and the retweet count was summed up using the individual retweet counts of all the tweets in that week. Manual annotations of relevant BLM-related events were added based on peaks in the retweet frequency.

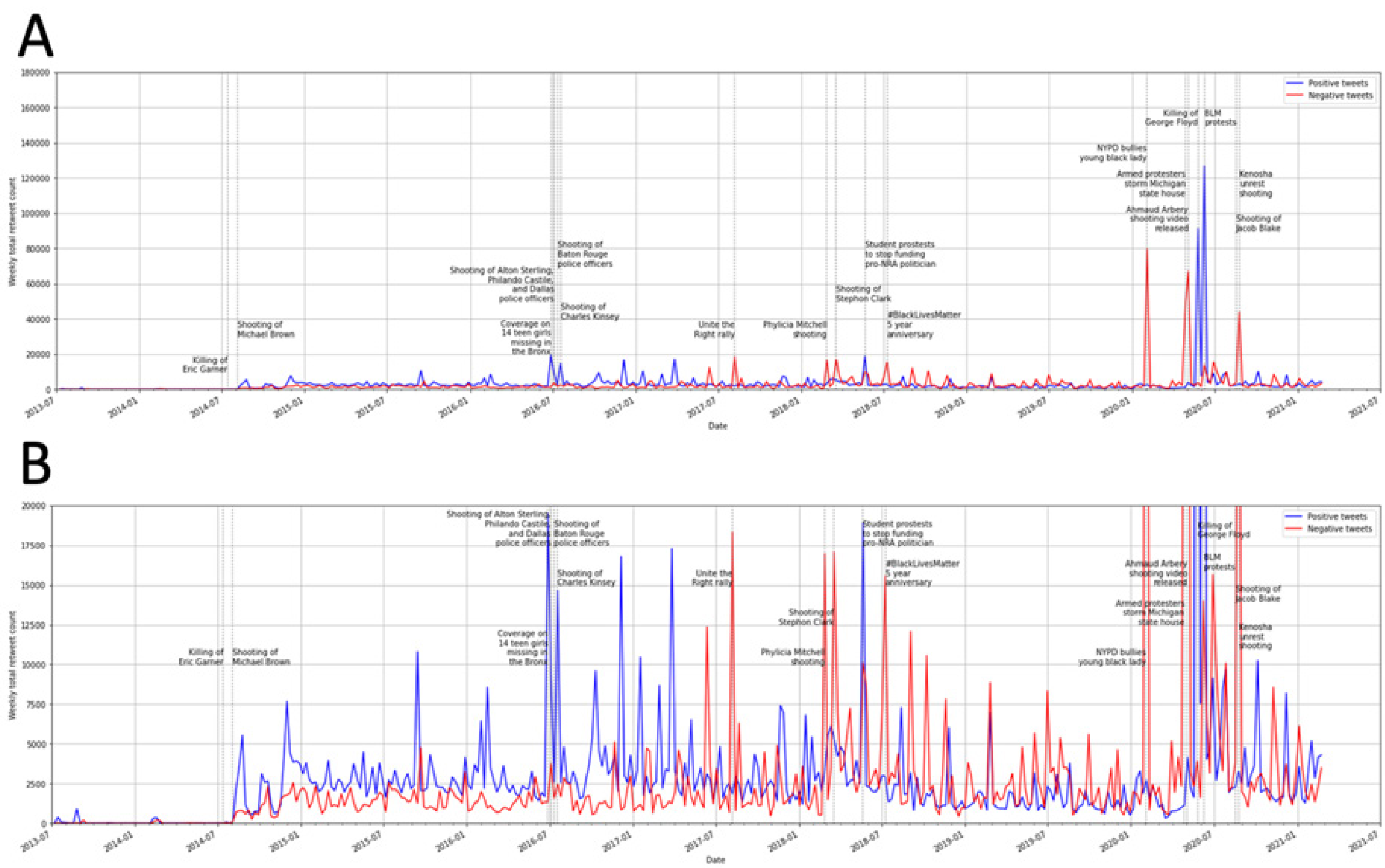


**Figure 7.** BLM Twitter corpus retweets over time with annotations for positive and negative tweets. Tweets in the BLM Twitter corpus predicted as having positive sentiment by the trained BERT sentiment analysis model with a confidence level greater than 0.8 were deemed "positive tweets" and tweets predicted as having negative sentiment with a confidence level greater than 0.8 were deemed "negative tweets". The method used to generate Figure 6 was used on both the corpus of positive tweets and the corpus of negative tweets to generate this plot of weekly retweet frequency over time from July 2013 to March 2021. The blue curve represents positive tweets, and the red curve represents negative tweets. B) is a zoomed in version of A) to better visualize smaller peaks.

enforcement system and advocate for the BLM movement. The spikes in the retweet frequency over time in the BLM Twitter corpus (Figure 6) is the result of people tweeting in protest to address the shootings and police brutality that systemically target the black community. The notable spike in 2014 corresponds to Eric Garner's death where an NYPD officer choked him to death on July 17, 2014. During this time, the phrase "'I can't breathe" was introduced, a phrase that has become a key part of the BLM movement [20]. The significant rise in the retweet frequency was again found in July 2016, as a result of the shooting deaths of Alton Sterling and Philando Castile. Both incidents were recorded and were subsequently released on social media. During the first three weeks of July, after the death of Alton Sterling, on average #BlackLivesMatter appeared more than 380,000 times a day on Twitter [2]. This is also consistent with the spikes found in the retweet frequency plot for 2016.

George Floyd's death on May 25, 2020 and the release of video evidence of his murder drew enormous attention and support to the BLM movement. Eleven days after his murder, the BLM protests peaked. Protests were held about 550 places across the US where between 15 to 26 million people participated physically and/or through social media [18]. The retweet frequency analysis also indicated the highest retweet frequency peak in June 2020 resulted from the nationwide BLM protests in justice for George Floyd. The second highest peak corresponds to the death of George Floyd on May 25. Retweets related to these two events surpassed all the retweet frequencies related to other BLM events indicating the tremendous impact of George Floyd's murder on the BLM movement.

In addition to analyzing retweet frequencies, the accompanying sentiment analysis of BLM tweets provided us with an understanding of public sentiment of the BLM movement throughout the different years. In general, positive sentiment was directed through tweets supporting the BLM movement, while negative sentiment was largely around tweets regarding the deaths and killing of black people. The movement started with many retweets of positive tweets supporting the BLM movement after it gained traction due to the killing of Eric Garner and Michael Brown. However, in 2018 many more negative peaks emerged, spurred by events such as the shooting of Stephon Clark, representing public anger towards police brutality and the killing of black individuals. In 2020, the BLM protests and activism following the killing of George Floyd received a large amount of support, as seen by the two large peaks in retweet frequency from positive tweets. Further analysis of the content of tweets is required to directly attribute fluctuations in sentiment to specific events.

Although sentiment analysis provides us with an overall picture of public sentiment towards movements such as BLM, it has limitations. One limitation is the difficulty of inferring, without context, a user's opinion from the sentiment of a tweet. For example, a user in favor of BLM could make a tweet mentioning BLM such as BLM needs to oppose black lives, enough is enough, which may be considred a negative statement tweet. Also, the classification of sentiment itself is nuanced; encouragement of activism over the deaths of black individuals can have positive sentiment because of the promotion of social justice or negative sentiment because of the mourning of lost lives or anger towards inequity. Tweets were either classified as positive or negative because the Sentiment140 dataset has binary labels (i.e., a tweet is either positive or negative). Other models may choose to classify a tweet as positive, negative, or neutral. This finer gradation allows for a more nuanced analysis, better considering the content of a tweet. We attempted to overcome the limitation of binary sentiment labels by only using the most confident positive and negative tweets for word frequency (Figure 5) and retweet frequency (Figure 7) analysis. However, we acknowledge the simplistic and restricted representation and prediction of sentiment throughout this study and the difficulty of using NLP to classify sentiment without additional real-world information and contextual clues.

In the future, beyond positive and negative sentiment labels, we can investigate other sentiments and emotions such as fear, anger, and joy in the BLM Twitter corpus. Additionally, the number of tweets in the BLM Twitter corpus was limited by Twitter's API tweet cap. Specifically, it returns a maximum of 500 tweets per query, and we ran our query on a daily timescale. While we assumed that having a sample of all relevant BLM-related tweets would be sufficient for our analysis, collecting the entire corpus would result in a more precise analysis. It would allow us to plot the frequency of sentiment over individual tweets rather than using the frequency of retweets, the latter of which assumes that the user who retweeted shares the same sentiment as the original tweet. However, to address this adequately, substantial computational resources would be required. While we only filtered for tweets in English, including tweets in other languages could be considered in the future to include the sentiments of a more diverse population of Twitter users.

## CONCLUSION

We used natural language processing and sentiment analysis to better understand the sentiment around, themes, and propagation of BLM movement, based on public sentiments expressed in Twitter data. We trained several machine learning models on a pre-annotated sentiment dataset, Sentiment140, to label tweets as having either positive or negative sentiment. After comparing the performance of the models using AUC-ROC as the metric, we found that the BERT model performed the best with an AUC of 0.97 on the training set and 0.94 on the testing set. The trained BERT model was utilized to recognize the sentiments of approximately 1 million tweets (dubbed the BLM Twitter corpus) from July 2013 to March 2021. From this, we found that words most frequently used in BLM-related tweets represented words describing activism, other social movements, racism, police, deaths, and political figures. Words describing activism and progressive political figures were related to positive sentiment, while words related to police and deaths were strongly associated with negative sentiment. We also found that there was a significant increase in retweet frequency around important events in the BLM movement, such as the killings of black individuals and related protests. Sentiment analysis revealed a large positive sentiment towards these protests, especially protests incited by the killing of George Floyd, and a large negative sentiment towards the shootings themselves. Therefore, Twitter sentiment analysis can shed light on the propagation of social-political movements, such as BLM, highlighting events that correlate with increased online attention to the movement and characterizing public sentiment around the movement over time. Researchers can further study why these events played an important role in the movement and activists can learn from these events to propagate the movement further and gain positive sentiment towards it.

## ACKNOWLEDGEMENTS

We would like to thank Jonathan André for reviewing this manuscript and for his insightful suggestions. JP, JSF, MM, AR, PM, and ARV initially designed the study. JP and ARV supervised the study. JP, SF, MM, AR, and PM trained and implemented the machine learning models. DO contributed to the revisions of the manuscript. All authors contributed to the writing of the manuscript.

## REFERENCES

1. Black Lives Matter. Herstory [Internet]. [cited 2022 Aug 30]. Available from: https://blacklivesmatter.com/herstory/
2. Anderson M, Toor S, Olmstead K, Rainie L, Smith A. 2. An analysis of #BlackLivesMatter and other Twitter hashtags related to political or social issues. Pew Research Center [Internet]. 2018 Jul 11 [cited 2022 Aug 30]. Available from: https://www.pewresearch.org/internet/2018/07/11/an-analysis-of-blacklivesmatter-and-other-twitter-hashtags-related-to-political-or-social-issues/
3. Hirschberg J, Manning CD. Advances in natural language processing. Science. 2015 Jul 17;349(6245):261–6. doi: https://doi.org/10.1126/science.aaa8685
4. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In: Proceedings of the ACL 2012 System Demonstrations [Internet]. Jeju Island, Korea: Association for Computational Linguistics; 2012 [cited 2022 Jan 30]. p. 115–20. Available from: https://aclanthology.org/P12-3020
5. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. JMIR Public Health Surveill. 2020 Nov 11;6(4):e21978. doi: https://doi.org/10.2196/21978
6. Cota DAM. Sentiment analysis of Black Lives Matter movement tweets. ResearchGate [Internet]. 2020 Jul 18 [cited 2022 Aug 30]. doi: http://dx.doi.org/10.13140/RG.2.2.18785.53603
7. Giorgi S, Guntuku SC, Rahman M, Himelein-Wachowiak M, Kwarteng A, Curtis B. Twitter corpus of the #BlackLivesMatter movement and counter protests: 2013 to 2020. ArXiv200900596 Cs [Internet]. 2020 Sep 28 [cited 2022 Jan 30]. doi: https://doi.org/10.48550/arXiv.2009.00596
8. Kumar S, Pranesh RR. TweetBLM: A hate speech dataset and analysis of Black Lives Matter-related microblogs on Twitter. ArXiv210812521 Cs [Internet]. 2021 Aug 27 [cited 2022 Jan 30]. doi: https://doi.org/10.48550/arXiv.2108.12521
9. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision [Internet]. 2009 [cited 2022 Aug 30]. Available from: https://

www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

10. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2022 Jan 30]. p. 785–94. doi: https://doi.org/10.1145/2939672.2939785
11. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi: https://doi.org/10.1162/neco.1997.9.8.1735
12. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics; 2019 [cited 2022 Jan 30]. p. 4171–86. Available from: https://aclanthology.org/N19-1423
13. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing [Review Article]. IEEE Comput Intell Mag. 2018 Aug;13(3):55–75. doi: https://doi.org/10.1109/MCI.2018.2840738
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2017 [cited 2022 Jan 30]. Available from: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
15. Lopez G. There are proven ways to keep protests peaceful. Trump is doing the opposite. Vox [Internet]. 2020 Oct 1 [cited 2022 Jan 30]. Available from: https://www.vox.com/2020/10/1/21427388/trump-protests-black-lives-matter-kenosha-wisconsin-portland-oregon
16. Wallach HM. Topic modeling: beyond bag-of-words. In: Proceedings of the 23rd international conference on Machine learning [Internet]. New York, NY, USA: Association for Computing Machinery; 2006 [cited 2022 Jan 30]. p. 977–84. (ICML '06). doi: https://doi.org/10.1145/1143844.1143967
17. Laurencin CT, Walker JM. A pandemic on a pandemic: Racism and COVID-19 in blacks. Cell Syst. 2020 Jul 22;11(1):9–10. doi: https://doi.org/10.1016/j.cels.2020.07.002
18. Drakulich K, Wozniak KH, Hagan J, Johnson D. Race and policing in the 2016 presidential election: Black lives matter, the police, and dog whistle politics. Criminology. 2020;58(2):370–402. doi: https://doi.org/10.1111/1745-9125.12239
19. Buchanan L, Bui Q, Patel JK. Black Lives Matter may be the largest movement in U.S. history. The New York Times [Internet]. 2020 Jul 3 [cited 2022 Jan 30]. Available from: https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html
20. Timeline: The Black Lives Matter movement. ABC News [Internet]. 2016 Jul 13 [cited 2022 Jan 30]. Available from: https://www.abc.net.au/news/2016-07-14/black-lives-matter-timeline/7585856
21. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011 Nov 1;12(85):2825–30. Available from: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html
22. Chollet et al. Keras [Internet]. 2015 [cited 2022 Aug 30]; Available from: https://keras.io
23. Wolf et al. Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations [Internet]. Association for Computational Linguistics; 2020 Oct [cited 2022 Aug 30]. p. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6
24. Abadi et al. TensorFlow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation [Internet]. USA: USENIX Association; 2016 Nov [cited 2022 Aug 30]. p. 265–283. Available from: https://dl.acm.org/doi/10.5555/3026877.3026899

# APPENDIX

## *Sentiment analysis models and model training details*

A logistic regression model can be used to model the probability of a binary dependent variable given a set of independent variables, also known as features, describing an observation. In the case of our study, the binary dependent variable was negative/positive sentiment of a tweet, and the features were based on the words in the tweet.

The equation for the logistic regression model is as follows:

$$p(x) = \frac{1}{1 + e^{-\beta \cdot x}}$$

where $p(x)$ represents the probability of the tweet being positive, $x$ represents the feature vector for the tweet, and $\beta$ represents a vector of coefficients learned by the model after training it on data. For the logistic regression model, features (words from a corpus) from the Sentiment140 dataset were extracted using a Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. Words were processed and assigned a score (or weight) on how relevant the word is, such that the classifier can filter out irrelevant words. The hyperparameters were tuned using a grid search over L2 regularization weights (0.1, 1.0, 10) with various solvers: liblinear, LBFGS, newton-cg, SAG, and SAGA. The final tuned logistic regression model used L2 penalty with a regularization weight of 1.0. No significant changes were found between the various solvers in the F1 and accuracy scores.

A naive Bayes classifier can also be used to model the probability of a binary dependent variable and, in our study, the probability of a tweet being positive. It uses Bayes theorem and the assumption that all features are independent. The equation for the naïve Bayes classifier is as follows:

$$p(C_{pos}|x) = \frac{p(C_{pos})p(x|C_{pos})}{p(x)}$$

where $C_{pos}$ represents the positive class (i.e. the tweet being classified as having positive sentiment) and $x$ represents the tweet as a feature vector.

XGBoost stands for eXtreme Gradient Boosting and is an algorithm/model that uses multiple decision trees in a way that allows for fast model training and optimized performance [10]. XGBoost has become popular recently owing to its success in solving machine learning problems. It can also be used to model the probability of a tweet having positive sentiment.

The naive Bayes classifier and XGBoost models also used the TF-IDF features created from the Sentiment140 corpus. One million words were chosen to train both models. The naive Bayes model was implemented using the MultinomialNB class in scikit-learn with the default parameters. For the XGB classifier, the final tuned model used the following hyperparameters: eta=0.9, min_child_weight=9, max_depth=9, max_leaf_nodes=1, gamma=9, max_delta_step=9.

Long short-term memory (LSTM) is a neural network architecture, specifically a recurrent neural network (RNN) architecture, that models sequential data and has the ability to learn long-term dependencies in the sequence data [11]. Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based neural network model and represents the state of the art for natural language processing tasks [12]. It is able to learn the contextual relationships between words in the text data. Both LSTM and BERT can also be used to model the probability of a tweet having positive sentiment.

The LSTM classifier has an input layer, a LSTM layer with 256 nodes followed by a ReLU activation function, and a final softmax output layer. Categorical cross entropy was used as the loss function and the ADAM optimizer was used for optimization.

The pre-trained BERT model was fine-tuned on the Sentiment140 training data. Tweets were pre-processed by being converted to lowercase, removing excess white spaces, replacing URLs with "URL", mentions like @username with 'AT_USER', and removing the '#' from hashtags. Tweets were then clipped or padded to have a maximum length of 128 characters to ensure all training data were of the same dimension. Training was done for two epochs with a batch size of 64. Cross entropy was used as the loss function and the ADAM optimizer, with a learning rate of 3e-5, epsilon value of 1e-8, and clipnorm value of 1, were used for optimization.