

Project Background and Goals

Dogwhistles: appeals of suggestive language that are meant to be only understood by a particular section of society, and thus contribute to an implicit kind of hate speech.



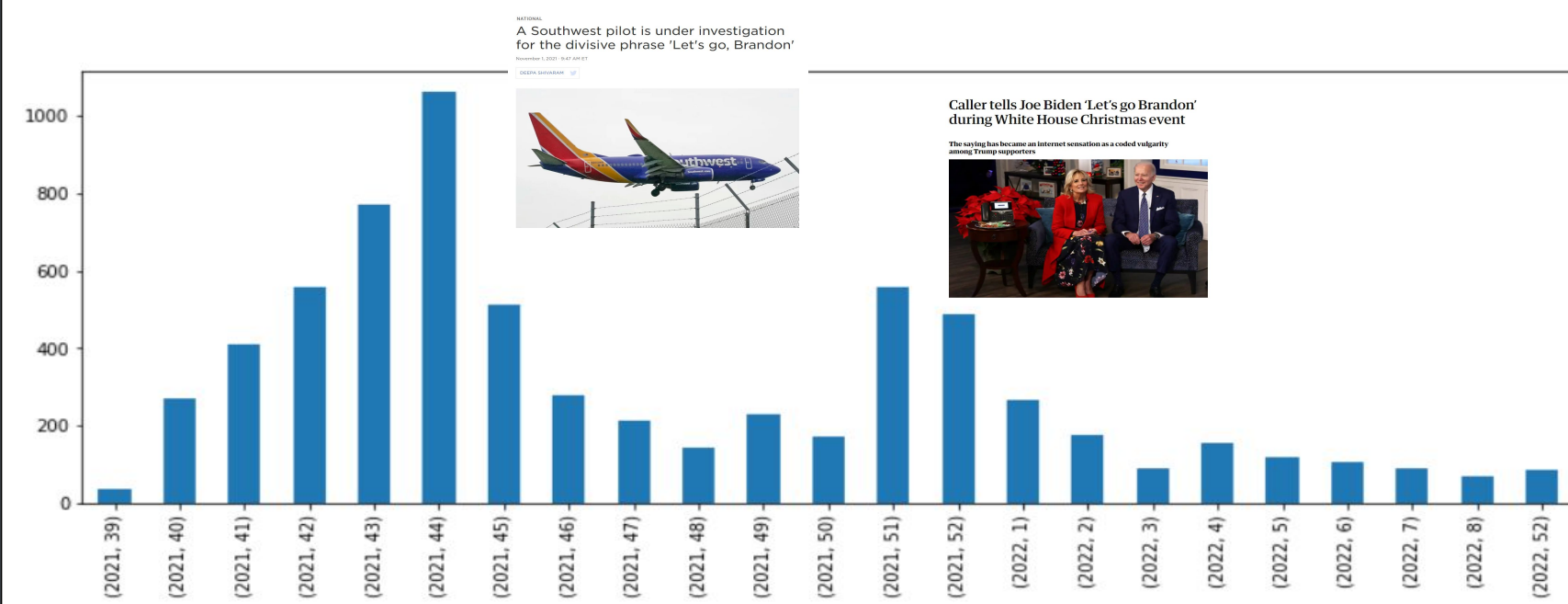
Case studies: "Let's go, Brandon!" (modern)
"Welfare Queen" (classical)

Guiding Research Questions:

1. Are there identifiable recurring themes in language using dogwhistles?
2. How do usage patterns of dogwhistles evolve over time?

Usage Over Time

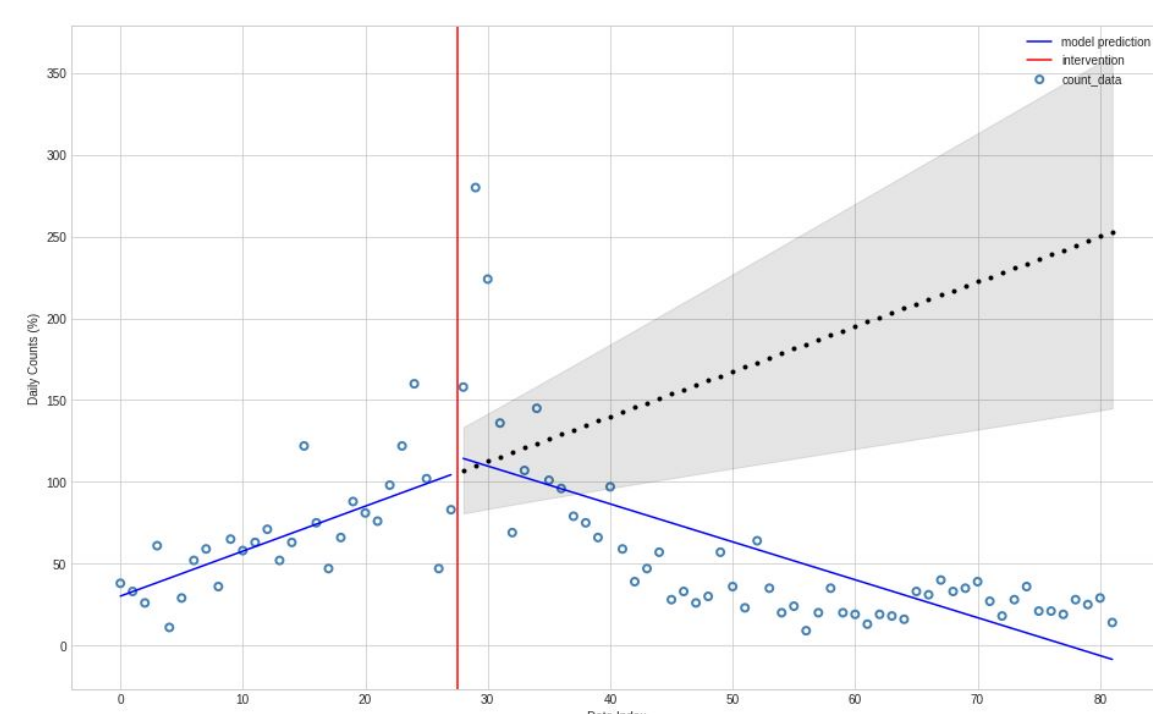
- Grouped dogwhistle occurrences by date to study longitudinal patterns.
- Conducted intervention tests to uncover intervention effects



Peaks in usage correspond to real life events where dogwhistle was used

Peak	Intervention Test	Coefficient	P-Value
P1	Regression	7.19	0.65
P1	ARIMA	38.24	0.054
P2	ARIMA	125.64	0.006

Peaks in usage correspond to real life events where dogwhistle was used



Data Collection

- Collected Twitter data using Twitter API v2 and Reddit data using Pushshift API.
- Collected Twitter data between 2.20.22 and 2.28.22
- Collected Reddit data between 07.2021 and 02.2022

Dataset	Count
Twitter "Let's go, Brandon!" Samples	6856
Reddit "Let's go, Brandon!" Samples	7178
Twitter "Welfare Queen" Samples	2943
Reddit "Welfare Queen" Samples	3406

Political Content Modeling



Categories	No. Tweets
External Relations	120
Freedom and Democracy	61
Political System	610
Economy	60
Welfare and Quality of Life	170
Fabric of Society	159
Social Groups	49

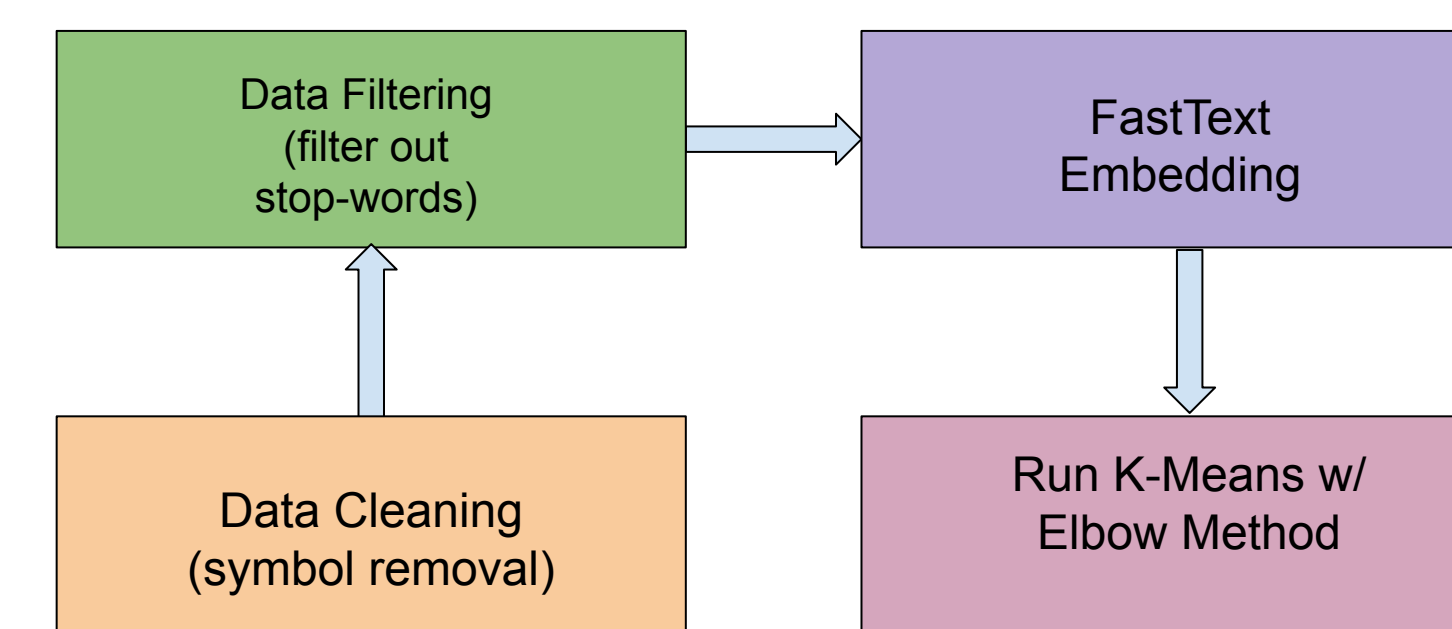
- Hand-verification of N=135 samples
 - model not transferable to shorter form text like Tweets

Sentiment Analysis



	Positive Samples	Negative Samples
Twitter	461	327
Reddit	74	25

Clustering



Clustering pipeline for Twitter and Reddit

Cluster (K=4)	Count
Anti-Biden	880
Unrelated "Brandons"	1015
Economy/Ukraine	2112
Spanish	145

Generated clusters from Twitter "Let's go, Brandon!" data

Cluster (K=4)	Representative Subreddits
Question Forums	r/AskReddit, r/TooAfraidToAsk
Conservative Politics	r/ConservativesOnly
News	r/TheNewsFeed, r/TheTwitterFeed
Miscellaneous	r/walkaway, r/WayOfTheBern

Generated subreddit clusters from Reddit "Let's go, Brandon!" data

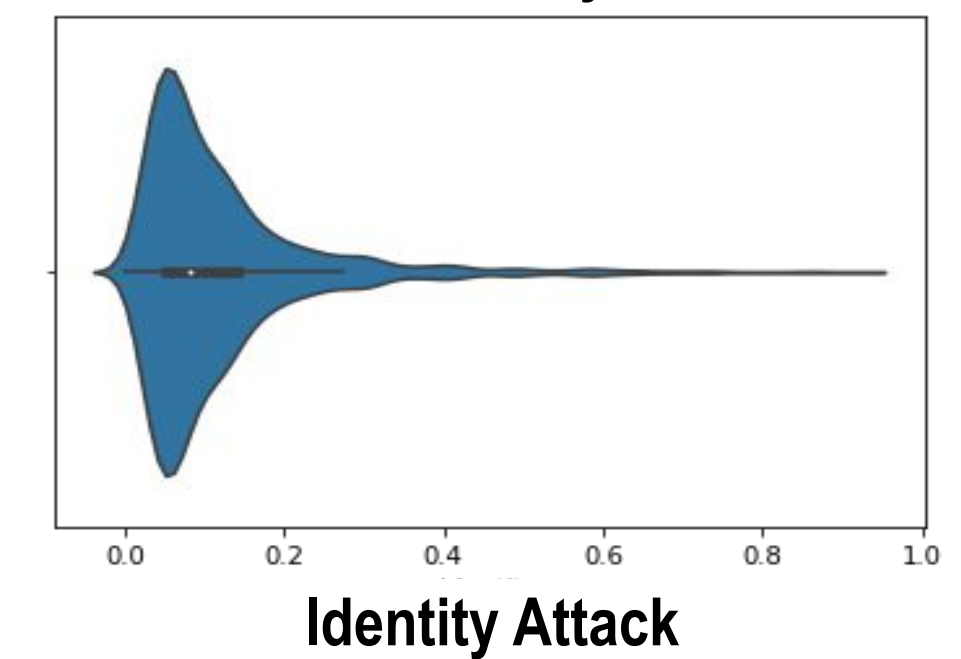
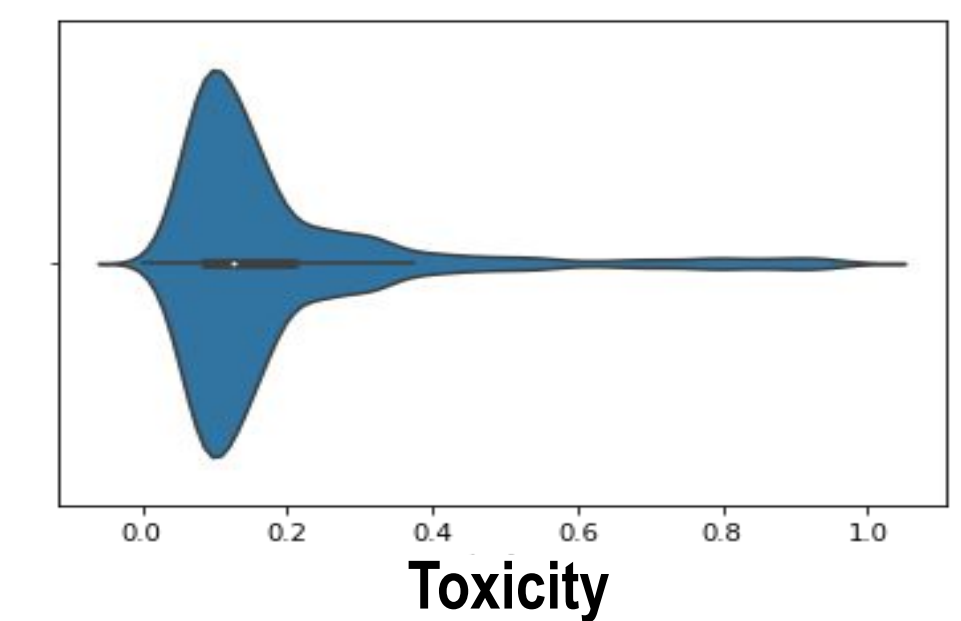
Cluster (K=5)	Count
White Nationalism and Racism	131
Anti-Government	71
Jeff Bezos	19
Personal Attack	116
General Noise	140

Generated clusters from Twitter "welfare queen" data

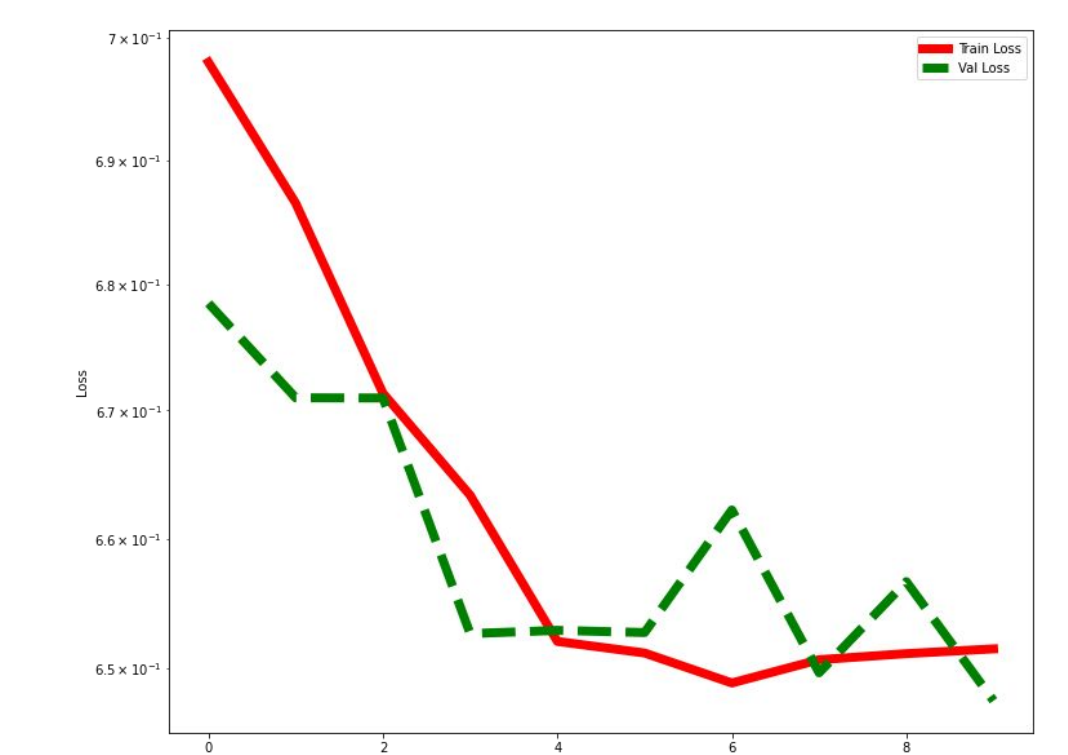
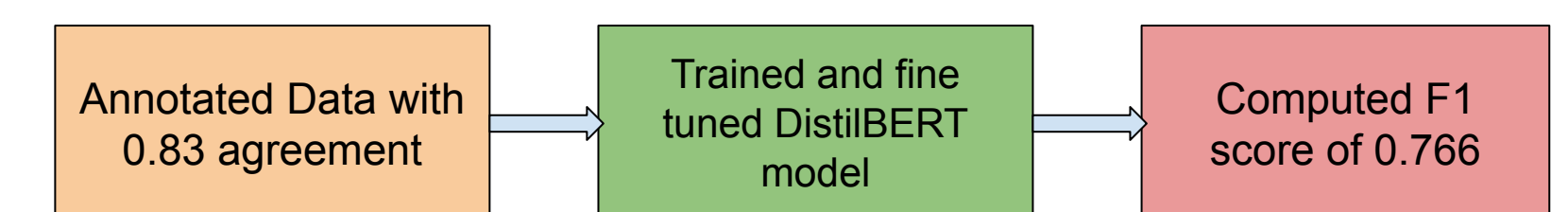
Toxicity and Personal Attacks

Perspective

- Used the perspective API to generate Toxicity and Identity Attack scores for data.
- Most data points have very low scores.



BERT Annotation Model



Loss curve for every epoch run

Future Work

- Leverage more user history-related features for context
- Experiment with alternative encoding strategies.
- Explore the generalizability and explainability of our model for more dogwhistles.

References

- Google Counter Abuse Technology Team, J. 2020. Perspective API. URL <https://www.perspectiveapi.com/how-it-works/>.
- Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; J'egou, H.; and Mikolov, T. 2016a. FastText.zip: Compressing text classification models. arXiv preprint arXiv:1612.03651.
- Laurer, M. 2020. Policy-DistilBERT. Unpublished paper.
- Heitmann, M.; Siebert, C.; Hartmann, J.; and Schamp, C. 2020. More than a feeling: Benchmarks for sentiment analysis accuracy.
- MacQueen, J.; et al. 1967. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, 281-297. Oakland, CA, USA.
- <https://developer.twitter.com/en/products/twitter-api>
- <https://github.com/pushshift/api>