

A Comparative Analysis of Semi-Supervised and Self Supervised Classification for Labeling Tweets about Police Brutality

Elaine Okanyene Nsoesie,
Wuraola Fisayo Oyewusi,
Opeyemi Osakuade,
Olubayo Adekanmbi



ABSTRACT

Social media has proven to be influential in social justice advocacy. In 2020, thousands of Nigerians and allies used the EndSARS hashtag to protest police brutality in Nigeria. In this work, we aim to understand the conversation associated with the EndSARS hashtag by comparing the outcome of semi-supervised and self-supervised machine learning classification algorithms for the automatic labeling of tweets. The self-supervised, zero-shot learning algorithm had the best performance for automatic tweet labeling with average weighted recall of 0.73, compared to cosine similarity with TF-IDF(0.71), cosine similarity with universal sentence encoder (0.58) and Jaccard Similarity(0.222). The major topics of discussion included complaints about police brutality events, Lekki Massacre, activism, media coverage, lack of response from public figures and questions about moving forward.

Introduction

Police brutality and EndSARS

In recent years, there have been a number of protests on the Internet and in-person demonstrations against police brutality around the world. The EndSARS social movement against police brutality which started in 2017 as a Twitter campaign using the hashtag #EndSARS to demand the disbanding of the unit by the Nigerian government, experienced a revitalization in early October 2020. On the night of 20 October 2020, at about 6:50 pm, members of the Nigerian Army were reported to have opened fire on peaceful EndSARS protesters at the Lekki toll gate in Lagos State, Nigeria. Amnesty International stated that at least 12 protesters were killed during the shooting.

In this work, we analysed tweets related to the EndSARS hashtags, we used unsupervised learning to cluster and perform topic modeling on the tweets to select label topics and present a comparison between semi-supervised and self-supervised classification for EndSARS tweets using semantic text similarity and Zero-shot classifier, respectively.

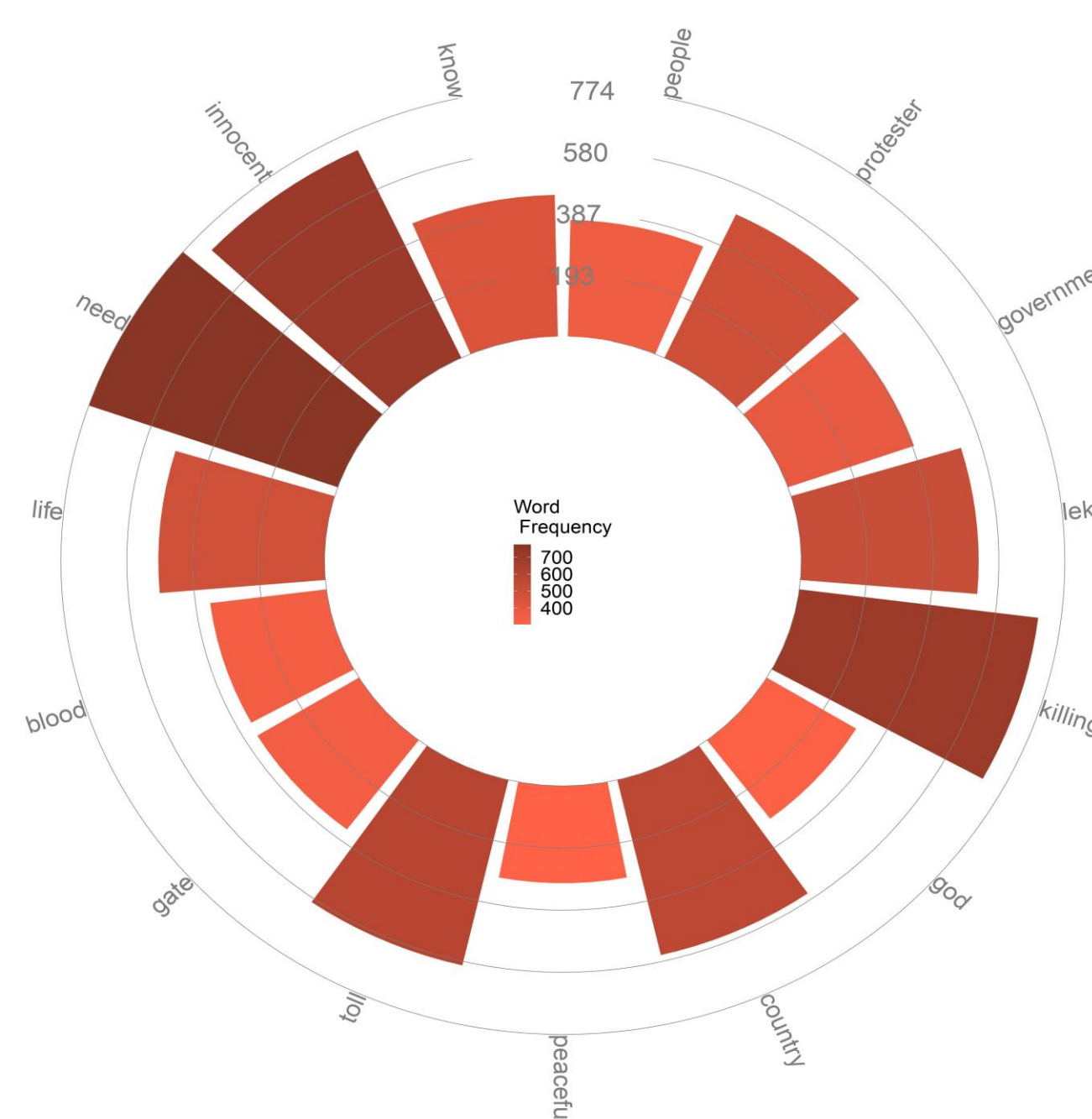
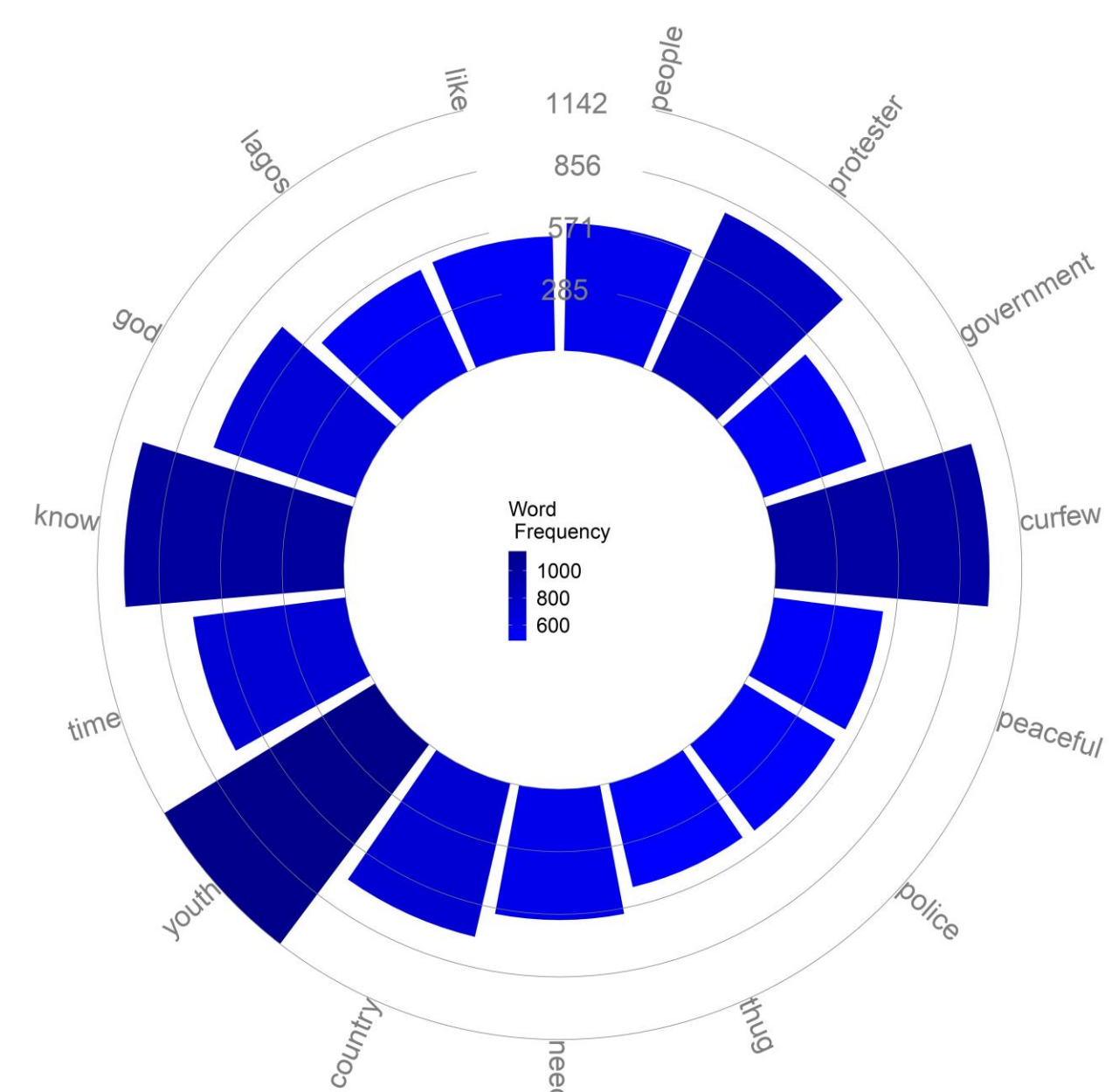
Methodology



332,371 tweets with EndSARS hashtags were collected between 1st of October, 2020 to 31st December, 2020 using the Twitter streaming API. We converted all words to lowercase letters and removed @user, symbols, links, Irrelevant hashtag related tweets, and stopwords

Exploratory Data Analysis and Label selection

We explored a portion of the dataset that focused on the Lekki massacre event, where Figure 2 shows the most frequent words before 6:50 pm on that day, which reveals tweets about curfew, government, peaceful, protesters etc. Figure 4, on the other hand, shows the word cloud for tweets from 6:50 pm till the end of the day, which reveals tweets about killings, lekki, and blood.



We explored topic modeling for the tweet clustering with LDA model and extracted 6 topic clusters, where each topic has different words that are interconnected.

The six topics extracted are as follows: Activism, Police brutality, Media coverage, Lekki massacre, Questions about moving forward/back, No response from public figures.

Results and Conclusion

Table below shows sample tweets and the predicted labels for each method, comparing each of the predicted labels to the human labels,

s/n	Sample tweet	Human label	USE label	Jaccard label	TF-IDF label	Zero-shot label
1	is there a missing person around you particularly during the movement friends and family should put out their information through the link below	Activism, Questions about moving forward/past, Media coverage	Media coverage, Lekki Massacre, Police brutality events	Lekki Massacre, Media coverage, Questions about moving forward/past	Police brutality events, Activism, Media coverage	Activism, Questions about moving forward/past, Media coverage
2	the change begins with you yes you	Activism	Activism, No Response from public figures, Questions about moving forward/past	Lekki Massacre, Media coverage, Questions about moving forward/past	Questions about moving forward/past, Activism, Irrelevant	Activism, Questions about moving forward/past, Media coverage

Zero-shot learning performed better especially with predicting accurate labels for the top 1 prediction. Although for tweets 2 and 3, USE labels were accurate for the top 1 prediction, the send labels are not related to the tweet, e.g., Lekki Massacre label should not be part of the top 3 labels for tweet 3 Jaccard similarity did not perform well on these tweets, since its predictions were almost the same for all tweets. Zero-shot learning proves to be a better approach to providing a more generalized label for multi-label tweets.

Method	Hamming Loss	Weighted Average Precision	Weighted Average Recall	Weighted Average F1 score
Cosine similarity with TF-IDF	0.349	0.42	0.71	0.50
Cosine similarity with USE	0.397	0.44	0.58	0.44
Jaccard Similarity	0.535	0.02	0.22	0.04
Zero-Shot Learning	0.343	0.53	0.73	0.51

References

- [1] Oscar Araque, Ganggao Zhu, and Carlos A Iglesias. A semantic similarity-based perspective of affect lexicons for sentiment analysis. Knowledge-Based Systems, 165:346–359, 2019.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. the Journal of machine Learning research, 3:993–1022, 2003.
- [3] Nina Cesare, Olubusola Oladeji, Kadija Ferryman, Derry Wijaya, Karen D Hendricks-Muñoz, Alyssa Ward, and Elaine O Nsoesie. Discussions of miscarriage and preterm births on twitter. Paediatric and perinatal epidemiology, 34(5):544–552, 2020.
- [4] Dumitru Erhan, Aaron Courville, Yoshua Bengio, and Pascal Vincent. Why does unsupervised pre-training help deep learning? In Proceedings of the thirteenth international conference on artificial intelligence and statistics, pages 201–208. JMLR Workshop and Conference Proceedings, 2010.
- [5] Herman Kamper, Aren Jansen, and Sharon Goldwater. A segmental framework for fully-unsupervised large-vocabulary speech recognition. Computer Speech & Language, 46:154–174, 2017

