

Research Article

Temidayo Michael Oladele, and Eniafe Festus Ayetiran*

Social Unrest Prediction Through Sentiment Analysis on Twitter Using Support Vector Machine: Experimental Study on Nigeria's #EndSARS

<https://doi.org/10.1515/opis-2022-0141>

received September 13, 2022; accepted February 17, 2023

Abstract: Social unrest is a powerful mode of expression and organized form of behavior involving civil disorders and acts of mass civil disobedience, among other behaviors. Nowadays, signs of most social unrest start from the social media websites, such as Twitter, Facebook, etc. In recent times, Nigeria has faced different forms of social unrest, including the popular #EndSARS, which began on Twitter with a demand that government disband the Special Anti-Robbery Squad (SARS), a unit under the Nigerian Police Force for alleged brutality. Mining public opinions such as this on social media can assist the government and other concerned organizations by serving as an early warning system. In this work, we collected user tweets with #EndSARS from Twitter and pre-processed and annotated them into positive and negative classes. A support vector classifier was then used for classifying the sentiment expressed in them. Experimental results show 90% accuracy, 94% precision, 85% recall, and 89% *F1* score on the test set. The codes and dataset are publicly available for research use.¹

Keywords: social unrest, support vector machine, Twitter, #EndSARS, sentiment classification

1 Introduction

The impact of social unrest on the peaceful existence and development of a society cannot be overemphasized. Social unrest in Nigeria is a serious social problem that affects the socioeconomic stability of the country. Social unrest also affects the progress and advancement of a nation. In Nigeria, whenever social unrest happens, it not only affects the socioeconomic stability but also the political stability of the country. Social unrest can stymie the economic growth and development of a country, lowering the international confidence level as well as increasing insecurity and the level of uncertainty of foreign investors. Osaghae (2015) highlighted the possibility of the country disintegrating in 15 years citing the report published by the Central Intelligence Agency in 2005 due to civil unrest, strikes, and violence. The researcher noted that social unrest is a menace whose impacts and costs cannot be truly quantified as it transcends just material

¹ <https://github.com/Temidayomichael/Social-unrest-prediction>.

* **Corresponding author: Eniafe Festus Ayetiran**, Department of Computer Science, Achievers University, Owo, Nigeria; Department of Computer Science, Norwegian University of Science & Technology, Trondheim, Norway, e-mail: eniafe.ayetiran@achievers.edu.ng, eniafe.ayetiran@ntnu.no
Temidayo Michael Oladele: Department of Computer Science, Achievers University, Owo, Nigeria, e-mail: dayoo.oladele@gmail.com

and financial costs to emotional trauma and possibly the loss of lives. Recent advances in technology and the consequent emergence of social media have undoubtedly offered humanity several innovative ways to communicate and share useful information regardless of location and other conventional constraints. Social media have been reported by several researchers to be one of the main tools leveraged in recent times to fuel social unrest globally. Chiamogu, Obikeze, Chiamogu, and Odikpo (2021) stated that there has been a massive growth in the rate of internet penetration as well as the production and consumption of social media content. They noted that this increase has greatly influenced participation in politics as well as policy formulation and governance in most countries of the world, especially in countries with emergent liberal democracies where populism has been misapplied. These countries, according to them, have duly minimized the conventional checks and balances existent in every democratic government, leading to the subversion of justice and unfortunate strangulation of the citizenry, who consequently turn to social media as a veritable platform for activism, justice-seeking awareness creation on socio-political issues, and to mobilize protests. Over the years, platforms such as Facebook, WhatsApp, Instagram, Twitter, and other social media platforms in Nigeria have radically altered how information is shared and disseminated. This development enables the seamless transmission of information about notable events, trends, and opinions, as well as promotes calls for accountability on government's actions and misdeeds. Chiamogu, Obikeze, Chiamogu, and Odikpo (2021) stated that social movements such as “#BringBackOurGirls,” “#RevolutionNow,” and “#EndSARS” gained traction by uniting people across the country and in the diaspora around a common cause through social media. The emergence and subsequent adoption of social media over the years in Nigeria have resulted in enormous growth in the amount and speed at which information is created, transmitted, and consumed. While this trend has ushered in several opportunities for socio-economic growth, social media has been a critical platform in recent times for disseminating hate speech, political activism, and the proliferation of fake, distorted, or exaggerated news, as well as inciting violence and civil unrest. Samantha and Olu (2010) reported that technology-savvy youths are using social media to raise funds in bitcoins, United States of America dollars, Nigerian nairas, and a variety of other currencies in support of the “#EndSARS” campaign, as well as mobilize support from celebrities, international media, and institutions both at home and abroad. Although the #EndSARS protest began peacefully, it was swiftly taken over by hoodlums and politically motivated groups, which led to quick escalation. This escalation further led to shooting of citizens, part of which is the infamous “Lekki massacre,” which the Nigerian government has denied its occurrence, despite international findings. While social media has become a veritable tool for demanding accountability in governance, exposing various ills and occurrences in society, and promoting products and services offered by businesses, it also has potential to cause social unrest and catastrophic breakdown of law and order as well. This has been a major source of concern in recent times. The masses can also easily be incited to act in civil disobedience over fake and unverified information disseminated by people with selfish interests. Audu (2018) noted that over the years, conflict along the ethnoreligious fault lines has been increasing at an alarming rate due to several factors, among which is the rapid rate at which unverified information is disseminated via social media. The author further noted that much more dangerous than the proliferation of fake news is the intentional exaggeration or distortion of true stories, as these types of issues are much more difficult to spot. To stem this rather ugly trend, which could lead to social unrest, several studies have been conducted on suitable means to detect and flag potentially chaotic social media posts and messages (Cadena et al., 2015; Compton et al., 2013; Ranganath et al., 2016). Recent advances in fields such as Artificial Intelligence, Data Science, Big Data, and Machine Learning have led to some solutions, allowing for more efficient social media data collation and analysis. Korolov et al. (2016b) noted that social media have grown to become a salient tool for protest. In recent times, one can hypothesize a relationship between the volumes of a particular type of social media communication. Korolov et al. (2016b) noted the presence of challenges in identifying relevant and genuine messages in a relatively large stream of messages. As such, one suitable method of solving this problem is the use of relevant data mining techniques to extract relevant data from Twitter and/or other social media platforms and employing Natural Language Processing (NLP) methods to analyze social media threads. In this work, we model public opinions on #EndSARS as a binary classification problem using data obtained from Twitter and annotated using standard procedure. The need to leverage recent advances in technology,

especially in the fields of data science and machine learning, to develop a suitable model for analyzing social media data for the likelihood of social unrest is timely. We have developed a model for the analysis of tweets with the capability of classifying them based on the potential of causing civil unrest or otherwise. The implication of this is that it will enable interested people, organizations, and relevant authorities to develop early warning systems to inform them about potential social unrests. The rest of the paper is organized as follows: Section 2 discusses the literature. In Section 3, the methodology including data development and experiments, is presented. Section 4 reports on the evaluation of the developed model, while Section 5 concludes the article.

2 Literature Review

Social media platforms such as Facebook and Twitter currently attract millions of users around the globe, giving them an open platform to share their thoughts and opinions (Mouthami et al., 2013). Traditionally, sentiments are classified in binary form, i.e., either positive or negative (Mouthami, Devi, & Bhaskaran, 2012) but have evolved over the years with the addition of the neutral class (Ayetiran, 2022) in numerous tasks. Sentiment classification is an offshoot text classification (Ayetiran, 2020; Novotný, Ayetiran, Stefánik, & Sojka, 2020; Sebastiani, 2002), including the ones with applications in the legal domain (Ayetiran, 2013, 2017). There are several machine learning methods for sentiment categorization such as Maximum Entropy, a feature-based model that does not take independent assumptions (Cheong & Lee, 2011). Stochastic Gradient Descent is another machine learning-based algorithm that is capable of making the classifier learn even if it is based on a non-differentiable loss function (Bifet & Frank, 2010). Breiman (2001) introduced a Random Forest, which targets the enhancement and storage of classification trees. Wang, Can, Kazemzadeh, Bar, and Narayanan (2012) introduced a system for real-time Twitter analysis with an experiment on the 2012 U.S. presidential election cycle. Naïve Bayes is another easy-to-implement and efficient machine learning algorithm and was originally proposed by Thomas Bayes (Husain & Singh, 2014). Support vector machine (SVM) is one of the most famous supervised machine learning-based classification algorithms (Khairnar & Kinikar, 2013). Different studies and research featuring these different machine learning algorithms and tools are available on sentiment analysis, but as compared to several other methods, SVM performed better in terms of accuracy and efficiency as the results were comparatively higher (Kumar & Sebastian, 2012; Pak & Paroubek, 2010). There are some research studies on social unrest in the fields of social and political sciences, each employing varying approaches to the subject. Studies have been carried out at the macro level: looking at the influence of the political and economic situation in general on protest moods (Dalton, Sickle, & Weldon, 2010) and viewing protests as organized social movements (Porta & Diani, 2007). From the perspective of an individual, scholarly research has been concerned with recruitment and activism (Verhulst & Walgrave, 2009; Walgrave & Wouters, 2014), as well as factors and cognitive processes affecting individual participation in protests (Klandermans & Oegema, 1987; Stekelenburg & Klandermans, 2013; Stekelenburg, 2013). Such factors have also been used for modeling protest participation (Klandermans & Oegema, 1987; Kelloway, Francis, Catano, & Teed, 2007; Stekelenburg & Klandermans, 2013) and in a recent study combining social science with a computational simulation of the dynamics of participants (Lu, 2016). Since online social media gained worldwide popularity, a wide stream of research studies have emerged on the extraction and structuring of the available information. Among the various social media websites, Twitter is especially popular among scholars due to the availability of large numbers of data that can be collected while the event is unfolding – a naturally occurring experiment. Some approaches to information extraction from Twitter are described by Kumar, Morstatter, and Liu, (2014) and Imran, Elbassuoni, Castillo, Diaz, and Meier, (2013). Scholarly works taking advantage of these technologies have been concerned with the public response to extreme events (Tyshchuk, Li, Ji, & Wallace, 2013; Tyshchuk, Wallace, Li, Ji, & Kase, 2014; Vieweg, 2012), modelling human behavior (Tyshchuk & Wallace, 2015; Korolov et al., 2015, 2016c), forecasting actions (Korolov et al., 2015, 2016a), attitudes (Magdy, Darwish, & Weber 2015), and even election results (Tumasjan, Sprenger, Sandner, & Welpe 2010). Research

studies on the role of social media in protests emerged after the Arab Spring when its contribution to the emergence of revolutions in Arab countries became apparent. Studies were undertaken on the role of social media in revolutions (Oh, Eom, & Rao 2015; Tufekci & Wilson, 2012), information dissemination during protests (Starbird & Palen, 2012; Valenzuela, 2013), and activism and support of social movements (Freelon, McIlwain, & Clark, 2016; González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011; Magdy et al., 2015). Despite the abundance of research studies on social media and protest, studies focused on the prediction modeling of a protest using data from social media are scarce. In one of the few articles on this (Filchenkov, Azarov, & Abramov, 2014), the authors compare the prediction of a protest with the prediction of election results. A study reported by Compton et al. (2013) presents a solution, relying on a constructed vocabulary from data. A recent work (Cadena et al., 2015) addresses this problem using influence cascades as a predictor of protest. This study, however, also relies on a constructed vocabulary. Ranganath et al. (2016) exploit the previous activities of a social media users to predict whether their next tweet will be related to a protest. The relationship between social media activities and protests during the Arab Spring was established in Steinert-Threlkeld, Mocanu, Vespignani, and Fowler (2015), using hashtags to filter relevant tweets. Our approach also uses a geographical and topical grouping of tweets. The notable difference in our approach is that we were able to develop a standard dataset on which a model was trained for classification of tweets into positive and negative classes with excellent performance in all evaluation metrics used. It was observed that previous works have either used baseline classification with two or three machine learning models. As mentioned earlier, our choice of SVM is due to its great classification accuracy based on evaluation studies available in the literature (Kumar & Sebastian, 2012; Pak & Paroubek, 2010).

3 Methodology

The classification/prediction system processing pipeline is presented in Figure 1.

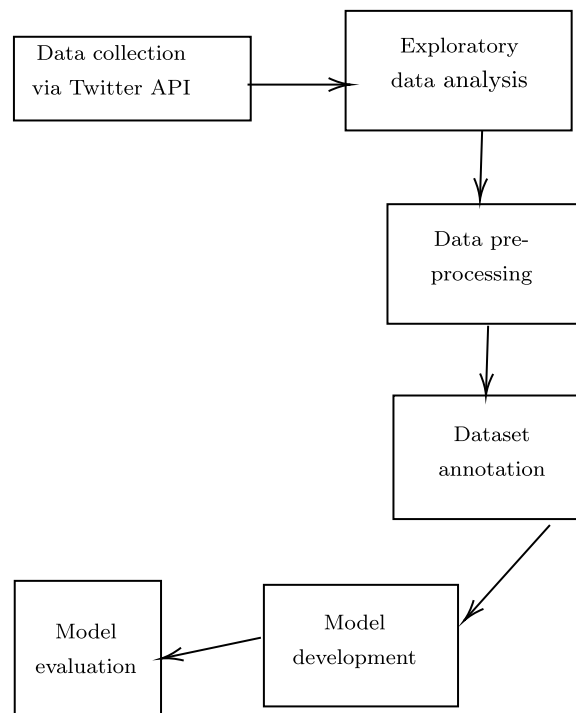


Figure 1: Prediction system development pipeline.

3.1 Data Analysis and Development

Twitter was searched for relevant tweets with #EndSARS using the Twitter Application Programming Interface (API). #EndSARS is a popular subject of conversation on Twitter, which started toward the commencement of the protest in October 2020 and continues till today. For the purpose of our experiment, 400 raw tweets were retrieved, with each consisting of several fields, including the tweet's (text), date, number of retweets (reposts), etc. Some of them also have null fields. These were discarded during the pre-processing phase.

3.1.1 Exploratory Data Analysis (EDA)

EDA was carried out on the selected dataset to help understand the dataset better and to help determine the pre-processing to be carried out on the data. The exploratory dataset carried out on the dataset includes data size and data description. EDA is also used to investigate and summarize the main characteristics of the dataset using visualization/plot methods. Diverse machine learning libraries were used through the Python programming language to carry out the exploratory analysis. Machine learning library is a library of functions and methods that enables easy scaling of scientific and numerical computations to streamline machine learning workflows. Under different emotional conditions, humans tend to behave differently. This includes the way they talk and express their feelings. Therefore, it is important to rely not only on the vocabularies used but also on the expressions and sentence structures used under the different conditions to quantify and model these feelings. Therefore, these were considered when manually selecting the 200 tweets. We used a Python library called Tweepy (Roesslein, 2020), developed to provide access to the official Twitter streaming Restful API with predefined methods that accept various parameters and return responses. The library is used to extract tweets containing one or more social unrest keywords for the purpose of this analysis.

3.1.2 Data Cleaning and Preprocessing

Since the dataset contains a lot of irrelevant and ambiguous information, preprocessing or data cleanup is crucial to prevent noise. As a result, before processing, extraneous values and other symbols that are not essential for classification purposes are deleted. The numerous preprocessing activities used in this investigation are as follows:

- (i) Removal of null data from the dataset.
- (ii) Investigating all elements within each feature.
- (iii) Cleaning of the dataset.
- (iv) Removal of stop words using the natural language toolkit (Bird, 2006).
- (v) Conversion of tweets to lowercase.
- (vi) Stemming of words.
- (vii) Punctuation removal.
- (viii) Count vectorization of the data.

3.1.3 Dataset Annotation

The preprocessed data are manually annotated by three expert annotators. Tweets on which they have a unanimous agreement were classified into two groups: positive and negative classes because almost all the collected tweets belong to these categories. A basic binary classification typically involves two classes, usually positive and negative for sentiment analysis. If the data are classified with positive and negative polarity, then the remaining data for which polarity cannot be decided due to disagreement among the annotators are earmarked for emoticon analysis. Furthermore, based on the emoticons used in the

earmarked data, tweets are classified accordingly as either having a positive or negative polarity. Thereafter, the remaining disputed tweets without emoticons are discarded. There are popular libraries used to achieve this like TextBlob (Loria, 2018), a text processing package for Python 2 and 3. It offers a basic API for doing standard NLP activities including part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and translation, among others. But the issue with these libraries is that they mainly detect emotions like sadness, happiness, etc. Furthermore, the results from these tools are not perfect and not good enough to entirely rely on in sensitive cases such as this. In our own case, we need to classify based on whether a tweet or text is in support of protest (positive) or not (negative). For reliability and accuracy, we develop the dataset manually based on unanimous agreement by three different annotators. Table 1 shows a sample of the annotated dataset with two categories: a class tagged as 0 (negative), which are the texts or tweets against social unrest, and a class tagged as 1 (positive), which consists of tweets in support of social unrest.

3.1.4 Emoticon Analysis

After the initial text classification phase, some of the data on which the annotators do not have a unanimous decision are discarded, but if they contain, emoticons are further analyzed. If the emoticon is classified as positive, the tweet sentiment is classified into a positive class appropriately and vice versa. The algorithm for classifying disputed tweets based on emoticon analysis works as follows:

- (i) The disputed tweets are searched for emoticon(s). If present, the emoticon(s) is/are analyzed for positive or negative emotion.
- (ii) Investigate the symbolic icon features in the emoticons. Each emoticon has a corresponding symbolic icon. For example, the symbolic icons for smiley and angry faces are :-) and >:(, respectively.
- (iii) Since emoticons can belong to either of two classes: positive, e.g., smiley, happy, and laughing, and negative, e.g., frown, angry, and sad. The emoticons are then classified accordingly by matching their symbolic representations to either of the two classes.

Samples of the manually annotated dataset are presented in Table 1 and count plot in Figure 2.

The statistics of the dataset are as follows:

- (1) 100 positive tweets from the selected overall 200 tweets supporting the social unrest campaign.
- (2) 100 negative tweets being the other half of the entire tweets that are against the protest.

3.2 Support Vector-based Classifier

SVM is one of the state-of-the-art text classification algorithms, which can be linear or non-linear. SVM aims to find a separating hyperplane (decision boundary) that is maximally far from any point in the training data. In this case, we apply linear (SVM) to a binary classification problem where a decision boundary is a

Table 1: Samples of the manually annotated dataset

Text	Label
The movement #EndSARS Now Is basically irrelevant pls... What we need now is to get them do their jobs not harassing people they are suppose to be protecting	0
We need them... SARS are special Anti-Robbery Squad... They're fighting Criminals in major cities...	0
Instead of the policemen to save this mans life, they stole from him as he was bleeding to death and helped the driver escape. Never forget! This is why we keep fighting!	1
This rubbish needs to stop. Shameless government and their pawns #EndSARS #end oppression https://t.co/4iflRBvYIM	1

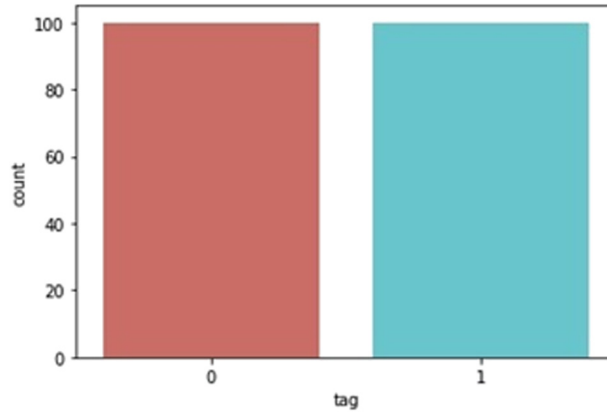


Figure 2: A count plot of the manually annotated dataset.

linear separator. The classification problem consists of a set of input vectors \vec{x}_i extracted from the tweets and y_i , the polarity of the tweets. Our linear binary classifier $f(\vec{x})$ is obtained using equation (1) as follows:

$$f(\vec{x}) = \text{sign}(\vec{w}^T \vec{x} + b), \quad (1)$$

where \vec{w}^T is the transpose of the weight vectors, sign denotes the decision based on the score $\vec{w}^T \vec{x} + b$. The two possible decisions are -1 and $+1$, which denote negative and positive polarities, respectively. The intercept b is used to determine the hyperplanes that are perpendicular to the normal vector.

3.3 Experiments

The dataset was split into train and test sets. SVM is used to train the model. In our experiment, training was done using 80% of the data, and testing was carried out on the remaining 20%. When the model has been fully trained, it is utilized for the model's final evaluation. Unlabeled tweets (test data) can be appropriately classified by the model whether they favor unrest or not based on the prior learning.

4 Model Evaluation

4.1 Evaluation Metrics

4.1.1 Confusion Matrix

A confusion matrix is one of the performance evaluation metrics of a classification-based machine learning model on a dataset where true values are known, thereby allowing the measurement of a model's accuracy, precision, recall and $F1$. A typical confusion matrix for binary classification is shown in Table 2.

Table 2: A Confusion matrix

		Predicted	
		False	True
Actual	False	True negative (TN)	False positive (FP)
	True	False negative (FN)	True positive (TP)

The basic measures for deriving the evaluation metrics are defined as follows:

- False positive (**FP**): Predicted as positive while the actual value is negative.
- True positive (**TP**): The actual value is positive and correctly predicted as positive.
- True negative (**TN**): The actual value is negative and correctly predicted as negative.
- False negative (**FN**): Predicted as negative while the actual value is positive.

4.1.2 Accuracy

Accuracy is the percentage of right predictions out of a total number of forecasts. That is, the percentage of the total number of correct predictions against the total number of predictions. It is defined by equation (2) as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (2)$$

4.1.3 Positive Predictive Value or Precision

Precision is the percentage of accurately detected positive cases, i.e., Precision is the actual, correctly classified data out of the overall classified data. It is defined by equation (3) as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (3)$$

4.1.4 Recall

Recall the percentage of true positive cases that are accurately detected. It is defined by equation (4) as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (4)$$

4.1.5 F1

F1 Score is the weighted harmonic mean of precision and recall is the F1. The closer the F1 score is near 1.0, the better the model's projected performance. F1 is obtained using the following equation:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5)$$

The model's performance is evaluated using confusion matrix, accuracy, precision, recall, and F1 score.

4.2 Results and Discussion

Experimental results obtained from the SVM classifier are shown in Table 3. The confusion matrix in Figure 3 shows 17 instances as true positives, 3 false negatives, 1 false positive, and 19 true negatives. The model

Table 3: Model prediction results

	Metrics			F1-score
	Accuracy	Precision	Recall	
Result	0.9000	0.9444	0.8500	0.8947

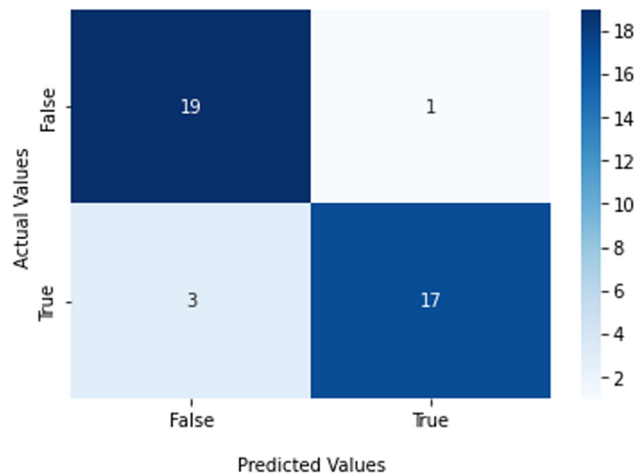


Figure 3: Model's confusion matrix with labels.

achieves a 90% accuracy, 94% precision, 85% recall, and 89% *F1* score. The results show that the model is better at accurately predicting negative instances. Observation on the test results shows that the total number of unique words and training data have a big influence on the accuracy rate attained. SVM has demonstrated to be a good classifier to learn these words with great predictive power when applied to a new dataset, leading to high accuracy in identifying sentiments in tweets.

5 Conclusion

Social unrest increases risks in the financial, economic, and political spheres, with possible long-term effects depending on the duration and frequency. An early warning system for these catastrophes may be helpful for policy makers, according to the rising frequency of such incidents over the previous few decades. Causes of unrest are diverse and interact with a wide variety of socioeconomic variables in complicated and non-linear ways, despite the fact that just a few indicators have been recognized in the literature as being significant in causing disturbance. To predict the occurrence of social unrest, in this work, we have annotated a dataset based on Nigeria's #EndSARS. Furthermore, we have utilized SVM to model public sentiments toward the protest. Experimental results show that the model generalizes well with good performance. This can be used in developing early warning systems for authorities about potential social unrest.

Acknowledgement: The authors immensely thank the three anonymous reviewers for their time and efforts, which significantly improved the initial manuscript. Part of this work was carried out during the tenure of the second author as an ERCIM “Alain Bensoussan” fellow.

Conflict of interest: The authors state no conflict of interest.

References

Audu, B. B. (2018). *How fake news in Nigeria compounds challenges to co-existence*. <https://institute.global/policy/how-fake-news-nigeria-compounds-challenges-co-existence>.

- Ayetiran, E. F. (2013). An intelligent hybrid approach for improving recall in electronic discovery. In M. Palmirani & G. Sartor, (Eds.), *Proceedings of the First JURIX Doctoral Consortium and Poster Sessions in Conjunction with the 26th International Conference on Legal Knowledge and Information Systems, JURIX 2013, Bologna, Italy, December 11–13, 2013*, vol. 1105 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Ayetiran, E. F. (2017). *A combined unsupervised technique for automatic classification in electronic discovery*. (PhD thesis), alma. Bologna, Italy: University of Bologna.
- Ayetiran, E. F. (2020). An index-based joint multilingual/cross-lingual text categorization using topic expansion via BabelNet. *Turkish Journal of Electrical Engineering and Computer Science*, 28(1), 224–237. <https://doi.org/10.3906/elk-1901-140>.
- Ayetiran, E. F. (2022). Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks. *Knowledge-Based Systems*, 252:109409. <https://doi.org/10.1016/j.knosys.2022.109409>.
- Bifet, A., & Frank, E. (2010). *Sentiment knowledge discovery in Twitter streaming data*. Conference Contribution.
- Bird, S. (17–21 July 2006). NLTK: The natural language toolkit. In N. Calzolari, C. Cardie, & P. Isabelle (Eds.), *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*. Sydney, Australia: The Association for Computer Linguistics. <https://doi.org/10.3115/1225403.1225421>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cadena, J., Korkmaz, G., Kuhlman, C., Marathe, A., Ramakrishnan, N., & Vullikanti, A. (2015). Forecasting social unrest using activity cascades. *PLoS One*, 10:e0128879. <https://doi.org/10.1371/journal.pone.0128879>.
- Cheong, M., & Lee, V. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, 13:45–59. <https://doi.org/10.1007/s10796-010-9273-x>.
- Chiamogu, A., Obikeze, O., Chiamogu, U., & Odikpo, E. (2021). Social media and group consciousness in Nigeria: Appraising the prevalence of socio-political protests. *Open Journal of Political Science*, 11:682–696. <https://doi.org/10.4236/ojps.2021.114043>.
- Compton, R., Lee, C., Lu, T.-C., Silva, L. D., & Macy, M. (2013). Detecting future social unrest in unprocessed Twitter data: Emerging phenomena and big data. *2013 IEEE International Conference on Intelligence and Security Informatics* (pp. 56–60).
- Dalton, R., Sickle, A., & Weldon, S. (2010). The individual–institutional nexus of protest behaviour. *British Journal of Political Science*, 40(1), 51–73. <https://doi.org/10.1017/S000712340999038X>.
- Devi, K. N., & Bhaskarn, V. M. (2012). Online forums hotspot prediction based on sentiment analysis. *Journal of Computer Science*, 8(8), 1219–1224. <https://doi.org/10.3844/jcssp.2012.1219.1224>.
- Filchenkov, A. A., Azarov, A. A., & Abramov, M. V. (2014). *What is more predictable in social media: Election outcome or protest action?* (pp. 157–161). Association for Computing Machinery. EGOSE '14: Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia (pp. 157–161). <https://doi.org/10.1145/2729104.2729135>.
- Freelon, D., McIlwain, C., & Clark, M. D. (2016). *Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice*. Social Science Research Network.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Sci Rep*, 1, 197. <https://doi.org/10.1038/srep00197>.
- Husain, M. S., & Singh, P. K. (2014). Methodological study of opinion mining and sentiment analysis techniques. *International Journal of Soft Computing (IJSC)*, 5:11–21. <https://doi.org/10.5121/ijsc.2014.5102>.
- Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. *Iscram*, 201(3), 791–801.
- Kelloway, E. K., Francis, L., Catano, V. M., & Teed, M. (2007). Predicting protest. *Basic and Applied Social Psychology*, 29(1), 13–22. <https://doi.org/10.1080/01973530701330884>.
- Khairnar, J., & Kinikar, M. (2013). Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, 3(6). <http://www.ijsrp.org/research-paper-0613.php?rp=P181364>.
- Klandermans, B., & Oegema, D. (1987). Potentials, networks, motivations, and barriers: Steps towards participation in social movements. *American Sociological Review*, 52(4), 519–531.
- R. Korolov, et al., “On predicting social unrest using social media,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, USA, pp. 89–95. 10.1109/ASONAM.2016.7752218.
- Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismael, M., & Wallace, W. (2016b). Predicting charitable donations using social media. *Social Network Analysis and Mining*, 6(1), 1–10.
- Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismael, M., & Wallace, W. A. (2016c). Predicting charitable donations using social media. *Social Network Analysis and Mining*, 6:1–10.
- Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismael, M., & Wallace, W. (2015). Actions are louder than words in social media. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 292–297).
- Kumar, A., & Sebastian, T. (2012). Sentiment analysis on Twitter. *International Journal of Computer Science Issues*, 9:372–378.
- Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics*. Springer Briefs in Computer Science.
- Loria, S. (2018). *textblob documentation*. Release 0.15, 2.

- Lu, P. (2016). Predicting peak of participants in collective action. *Applied Mathematics and Computation*, 274:318–330. <https://doi.org/10.1016/j.amc.2015.11.015>.
- Magdy, W., Darwish, K., & Weber, I. (2015). #Failedrevolutions: Using Twitter to study the antecedents of ISIS support. *First Monday*, 21(2). <https://doi.org/10.5210/fm.v21i2.6372>.
- Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013). *Sentiment analysis and classification based on textual reviews*. In *2013 International Conference on Information Communication and Embedded Systems (ICICES)* (pp. 271–276).
- Novotný, V., Ayetiran, E. F., Stefánik, M., & Sojka, P. (2020). Text classification with word embedding regularization and soft similarity measure. *CoRR*, abs/2003.05019.
- Oh, O., Eom, C., & Rao, H. R. (2015). Research note: Role of social media in social change: An analysis of collective sense making during the 2011 Egypt revolution. *Information Systems Research*, 26(1), 210–223.
- Osaghae, V. (2015). Causes of Nigeria unrest and conflict situation. *World Academy of Science, Engineering and Technology, International Journal of Humanities and Social Sciences*, 1:38–40.
- Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf.
- Porta, D. D., & Diani, M. (2007). *Social movements: An introduction*. Malden, MA, USA; Oxford, UK; Victoria: Australia: Blackwell Publishing.
- Ruppel, S., & Arowobusoye, O. (2010). *EndSars: How social media challenges governance – the case of Nigeria – PRIF Blog*. <https://blog.prif.org/2020/10/27/endsars-how-social-media-challenges-governance-the-case-of-nigeria/>.
- Ranganath, S., Morstatter, F., Hu, X., Tang, J., Wang, S., & Liu, H. (2016). Predicting online protest participation of social media users. In D. Schuurmans & M. P. Wellman (Eds.), *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12–17, 2016* (pp. 208–214). Arizona, USA: Phoenix, AAAI Press.
- Roesslein, J. (2020). *Tweepy: Twitter for Python!* URL:<https://github.com/tweepy/tweepy>.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>.
- Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted?: Information diffusion and the 2011 Egyptian uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*.
- Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., & Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, 4(1), 19. <https://doi.org/10.1140/epjds/s13688-015-0056-y>.
- Stekelenburg, J. (2013). The political psychology of protest. *European Psychologist*, 18:224. <https://doi.org/10.1027/1016-9040/a000156>.
- Stekelenburg, J., & Klandermans, B. (2013). The social psychology of protest. *Current Sociology*, 61:886–905. <https://doi.org/10.1177/0011392113479314>.
- Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir square. *Journal of Communication*, 62(2), 363–379. <https://doi.org/10.1111/j.1460-2466.2012.01629.x>.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welp, I. M. (2010). *Predicting elections with Twitter: What 140 characters reveal about political sentiment*. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), 178–185. <https://doi.org/10.1609/icwsm.v4i1.14009>.
- Tyshchuk, Y., Li, H., Ji, H., & Wallace, W. A. (2013). Evolution of communities on Twitter and the role of their leaders during emergencies. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM'13* (pp. 727–733). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2492517.2492657>.
- Tyshchuk, Y., & Wallace, W. (2015). *Modeling human behavior in the context of social media during extreme events caused by natural hazards*. Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute. PhD thesis, <https://dspace.rpi.edu/handle/20.500.13015/1446>.
- Tyshchuk, Y., Wallace, W. A., Li, H., Ji, H., & Kase, S. E. (2014). The nature of communications and emerging communities on Twitter following the 2013 Syria sarin gas attacks. In *Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference, JISIC'14* (pp. 41–47). USA: IEEE Computer Society. <https://doi.org/10.1109/JISIC.2014.16>.
- Valenzuela, S. (2013). Unpacking the use of social media for protest behavior. *American Behavioral Scientist*, 57:920–942. <https://doi.org/10.1177/0002764213479375>.
- Verhulst, J., & Walgrave, S. (2009). The first time is the hardest? A cross-national and cross-issue comparison of first-time protest participants. *Political Behavior*, 31:455–484. <https://doi.org/10.1007/s11109-008-9083-8>.
- Vieweg, S. E. (2012). *Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications*. (Doctoral dissertation), University of Colorado at Boulder.
- Walgrave, S., & Wouters, R. (2014). The missing link in the diffusion of protest: Asking others. *American Journal of Sociology*, 119(6), 1670–1709.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In *Proceedings of the ACL 2012 System Demonstration* (pp. 115–120). Jeju Island, Korea: Association for Computational Linguistics.