Stressor Classification of Filipino Political Tweets Using LDA, SVM, XGBoost, Logistic Regression

Mark Gabriel E. Edaño, Ryan JosephS. Gonzales,
Raphael Carlo B. Laguda and Joel C. De Goma
School of Information Technology
Mapua University - Makati
Makati City, Philippines
mgeedano@gmail.com, rjsgonzales21@gmail.com,
raphaellaguda99@gmail.com, jcdegoma@mapua.edu.ph

Abstract

With the advancement of technology, Filipinos have a means of connecting to social media mainly to share what they are doing or what they feel now. This could lead to people venting out their stress on platforms such as Twitter. One of the topics that cause people a lot of stress is Politics and many social media users share their opinions in a stressful manner on Twitter. This paper will focus on detecting the reason for stress called stressors from the tweet. This will be done by collecting tweets based on their hashtags and NLP technique called topic modeling specifically LDA to form the topic of stress for stress detection. Then Machine learning algorithms of SVM, XGBoost, and Logistic Regression will be used on the tweets and topics created by the Topic modeling to create and train a model that can predict stressors based on the tweets and topics.

Keywords

Political Tweets, Stressors, LDA, Machine Learning Algorithms, Stress

1. Introduction

1.1 Stress

Stress is a form of mental illness that people nowadays suffer from (Guntuku et al. 2019). With the advancement of technology, Filipinos have a means of connecting to social media mainly to share what they are doing or what they feel now. This could lead to people venting out on platforms such as Twitter. Stress affects a person's mental and at times physical health though for better or worse some people might have better mitigation to cope with stress. In other words, performance would be greatly affected if a person is suffering from stress. Based on Seaward (2018), stress should be managed and monitored, most people especially people from adulthood are not aware that they are under heavy stress.

1.2 Stress Detection

One way of detecting stress from social media is using an NLP technique called Topic Modelling in which collected social media post text are collected and put through a process called Latent Dirichlet allocation. In this process the top words from the collected post will be obtained and based on the top words the researcher can create their very own stressors.

Pillai et al. (2018) uses this topic Modelling Method to form 5 stressors. He used this on two different types of group tweets which are Political tweets and tweets about transportation. He used a tool to know which tweets are stressed and also manually annotated the tweets themselves. The tweets which were considered stress were the only ones used to topic modeling to create the topics.

1.3 The Problem, Gap, or Opportunity

This study hopes to contribute to the previous study (Pillai et al. 2018) a s some sort of a continuation with only stressors as the focus of detection. The discovered models can also be used by computer scientists who would like to specialize in NLP methods as well as future researchers who would to study and explore the possible stressors on tweets of Filipinos. This study will also produce a library of commonly used Tagalog words that would define a stressor as well as produce a dataset with annotated reasons for stress.

This study would only focus on the social media platform called Twitter and would solely focus on NLP methods to analyze tweets from Filipinos since Twitter is used by most Filipinos (Van der Schuur et al. 2018). The model would center around the predictions of stressors. On the other hand, although similar studies included relaxation (Pillai et al. 2018) in their method this study would be excluding that factor and would solely focus on stress and the possible stressor. The data-mined tweets will only be about politics since the majority of stress-related tweets are about politics (Pillai et al. 2018).

1.4 Objectives

Social media networking sites are commonly used by people to share their everyday details with the world. In this paper, we would determine the stressors on tweets of Filipinos with the challenges of the current lexicon tools not being well suited for scenarios with grammatical problems.

Thus, with the platform as the bridge for the study, the researchers would like to aim: To identify the stressors of Filipino Tweets from the Twitter platform using topic modeling methods, to find which model from the three machine learning algorithms provides the highest performance in terms of fi measure and accuracy and to discover what is the commonly used word that identifies the stressors.

2. Literature Review

2.1 Stress

Stress is one of the most common things that humans will ever experience in their lifetime. Inevitably there will come a time wherein we as social beings would encounter social media. Based on Seaward (2018) suggests that stress should be managed carefully because even though it is a regular feeling of emotion that we experience most people do not know that they are experiencing heavy stress. Almost everybody is using social media and a great majority of them are addicted to it so much that they become stressed and have a lack of sleep because of it which can be deemed harmful to us. Mental health conditions can be monitored by their language on social media (Guntuku et al. 2019).

2.2 NLP Studies on Social – Media

Social media posts are also one of the most common things to use NLP methods over. Hussein et al. (2018) made a study survey on NLP techniques in which he concluded that NLP methods on social media are a must in this day of age. The reason for this is that understanding a post in social media cannot be fully understood by reading the post itself that is why NLP techniques such as Topic Modelling analysis are important. Social media sites such as Facebook are very well known to have posted regarding how they feel. The study of Wang et al. (2020) and Hussein et al. (2018) focuses on how people behave, and they speculate that they might post similar posts, and within this type of post, they can determine the emotion of it using sentiment analysis.

2.3 Identification of Stress NLP methods

Social media is one of the most common ways to express your feelings in a text post. These text posts can be a post of stress expression. Pillai et al. (2018) focus on knowing whether a tweet is a stress state or a relaxed state. They used a TensiStrength as the basis of the range of stress or relaxation based on the tweets. The researchers mainly focus on collecting a bunch of tweets and not finding out the source of stress that came from the tweets, and they also do not have a focus group on the collection of tweets.

2.3 Code Switching

In the Philippines, the official languages are English and Filipino because this code-switching always happens. Code-switching is a prevalent phenomenon in multilingual communities in which people's conversations will alternate between two or more languages (Jose et al. 2020). This is also the same in tweeting, where Filipinos often tweet in a code-switch manner alternating between Filipino languages like Tagalog. Cebuano etc., and English (Abastillas et al.

2018). Since Twitter is a social media globally used, most tweets shown and collected in multilingual places have the structure of code-switching (Rijhwani et al.2017).

2.5 Topic Modelling

Topic Modelling is an NLP method where statistical modeling is used for finding topics in a collection of documents. LDA is one of the methods of topic modeling. LDA is, also known as Latent Dirichlet allocations, is used to classify text in a document to a particular subject (Gollapalli et al. 2018). LDA works by first knowing the numbers of words in a document, and then it will be put on a fixed set of topics, then the topics will be selected based on the multinomial distribution. Based on the collection of words on the multinomial distribution, a topic can now be formed based on those words. Topic modeling can also be used on a group of tweets because tweets can be considered documents to make a topic out of Negara et al. (2019). Topic Modelling can also be paired with a machine learning classifier algorithm to make a classifying model.

2.6 Political hashtags

Armstrong (2020) study focuses on the mass and the gossip around politics. The researcher stated that the government utilizes rumors and gossip to distract people from any problems in society. This could also be used as a unified front and a cultural bond to rally their right to speak. As has been found by other researchers, people are easily influenced by gossip. Even though it is often untrue, it spreads like wildfire using the media and leading to the people's view revolving around the elections, specifically past issues among the candidates. There are also cases where news is the leading front of how the people currently view the government. The study (Lintao and De Leon 2021) says politics on social media are just means of generating new ideas, reaffirming existing political beliefs, entertaining, and educating people.

2.7 Framework

As shown in figure 1, the process starts by collecting a lot of tweets and having enough to get data. Then after the data is collected, the pre-processing stage is commenced. The most common way of this process is using necessary techniques like tokenization, stemming, removing of stop words. When it comes to Twitter, the pre-processing stage can still be improved by removing special characters, URLs, numbers, punctuations, etc., and retweets. After the pre-processing techniques, the advice of a domain expert will be needed to label this tweet for them to identify which of the data can be considered a stressor. The dataset is split into different ratios for model training and testing, most commonly 70% and 30% ratios.

The Model training is done by having a classification method. It will be using different classification algorithms using SVM XGboost and Logistic Regression to know which classification algorithms would produce the best model with the highest accuracy.

Model Testing will then be done on the remaining percentage of the dataset to know if the method is consistent and to know which model classifier has the best accuracy on producing a result. The Evaluation is the last process in which the chosen Model will be used upon different data sets.

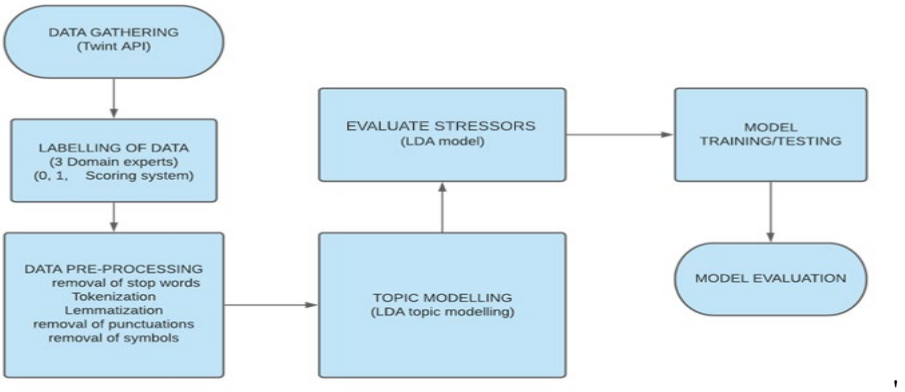

Figure 1. The framework of Stressor detection of Tweets

2. Methodology

2.1 Data Gathering

The study will mainly focus on collecting at least 6500 tweets from Twitter, the tweets will only consist of those from the area of the Philippines. The gathered tweets contain both English and Filipino language and will be separated into two datasets and were mined through TWINT.

The use of Twitter API will be the main tool used for the collection of tweets. The researcher will primarily focus on the Filipino Political tweets with hashtags such as #duterte, #dilawan, #BBMIsMyPresident2022, #PHVotes, #DOH, #Presidency, #COVID19 to focus the people's POV on the politics in which must be eventful to possibly gather more stressed tweets.

2.2 Labeling of Data (Determining if stressed or not)

The assistance of a domain expert is needed for the dataset to be completed. In the study, the researchers hired a psychometrician to know which data tweet is to be considered as stress. The Lazarus Stress theory (Cooper and Quick 2017) and Labelling theory (Sjöström et al. 2017) will be the criteria that the domain expert used to know if the tweet is stressed or not. The Labeling theory is also used by the Domain Expert because the nature of tweets being a text in a text word association must be put into place to know the nature of the tweet being stressed or not. Sample of labeling is shown in Table 1.

The Domain expert will annotate at least 3500 - 5000 tweets with scores of (0 -not stressed, 1 – stressed) to have enough annotated data to be used for topic modeling for finding potential stressors.

Table 1. Labeling of Dataset

Labeling of tweets			
Tweet	**DE1**	**DE2**	**DE3**
DUTERTE AT MARCOS PAREHONG TAKOT SA DALUYONG NG MAMAMAYANG LUMALABAN!! Parehong dinarahas ang mga nagpoprotesta nang payapa. Pareho ring isinusuka ng mamamayan. #NeverAgain #OustDuterte	1	1	1
nagkakaisa ang malawak na hanay ng sambayanan #oustduterte #wakasanna!	1	1	1

2.3 Data Pre-processing

Before pre-processing the dataset, the researchers first combined the datasets for each respective category (English and Tagalog). Through majority voting, the scores are then finalized with the ff example. As shown in Table 2, the majority will determine the label.

Table 2. Majority voting

	DE1	DE2	DE3	**Result**
Score	0	1	0	**0**
	0	0	1	**0**
	1	1	0	**1**

This process will result in a unified dataset of the domain experts for both the English and Tagalog dataset. The researchers then dropped all the tweets that were scored 0 or "not stressed" by the Domain Experts the reason being on to the next step, the words that are supposed to be clustered must be stressed only to have better results as to having a non-stressed tweet will make the clustered words possibly mixing in words that are not associated with stress. The resulting rows of the Tagalog dataset number 2,354 and as for the English dataset, 3,207. An example of pre-processed tweet is shown in table 3.

The following pre-processing is also proposed to be done for the tweets to be more accurate data to use (Negara et al. 2017):

- Removal of URLs since links are not a source of data the researchers are looking for.
- Retweets are ignored since retweets are not considered personal tweets from the user.
- Twitter unique symbols/letters are ignored (Hashtags, "@ username")
- All tweets are converted to lowercase
- Removal of stop words such as "are, as, a, am, etc." for both English and Tagalog words since those do not pose any significance on a tweet.
- Tokenization - each phrase/word is referred to as a token.
- Lemmatization - A process wherein a word/token is reduced to its word stem such as its roots.

Table 3. Pre-Processed Tweet

Original Tweet	Pre-Processed Tweet
Marcos, Duterte, walang pinag-iba! Parehong tuta, diktador, pasista! #NeverAgain #MarcosNotAHero #DuterteWakasan #OustDuterte	marcos duterte pinag parehong tuta diktador pasista neveragain marcosnotahero dutertewakasan oustduterte

2.4 Finding Stressors

The next step after preprocessing the dataset is to build an LDA topic modeling method. After creating a model of LDA, the researchers then looked for the dominant topics or the most used words per topic. After observing the top words per topic, the researchers again ask for the aid of the domain expert to evaluate the words per topic to come up with a general topic to represent the top used words in which the researchers will consider as the stressors. As shown in table 4, the five stressors that were formed are Political Stance, Government Policies, Election, Filipino Political News, President for the Filipino dataset and Election, Government Policies, Recollection, Filipino Political News, Political Stance for the English dataset in which is visualized below.

Table 4. Clusters and stressors for Tagalog and English Dataset

Most Common Words per Topic (Tagalog)	**Stressor**	**Most Common Words per Topic (English)**	**Stressor**
address, say, testing, kayo, people, ask, face, think, lawyere, class, test, pilipina, start, supporter, doh, apologist	Political Stance	respect, vote, country, dilawan, use, test, run, health, help, feel, covid, follow, leni, ask, know, leader	Election
Duterte, government, election, drug, people, corruption, administration, run, president, country, support, covid, try, year, allege, kill	Government Policies	presidency, kayo, election, face, hope, case, file, problem, make, drug, year, come, war, shield, doh, class	Government Policies
dilawan, neverforget, never again, bbm, law, country, aquino, presidency, forget, bbmforpresident, endbayan, hope, budget, politic, dictatorship	President	duterte, say, candidate, support, need, try, think, muna, oust, opposition, philippine, supporter, thank, corruption, work	Political Stance

After the evaluation of the Domain Experts on the cluster of words, we then asked the Domain experts to then label the dataset with the use of the evaluated stressors using the initial dataset before the preprocessing method. A sample labelling is shown in table 5.

Table 5. Labeled Dataset

Tweet from Dataset	**Stressor**
I finally registered to vote! Add 1 more ballot to #OustDuterte	Election
this administration has never prioritized the masses, only implementing self-serving neoliberal policies like the Build! Build! Build! program. #NoToCarbonPrivatization #EndStateFascism #DefundThePolice #OustDuterte	Government Policies

when I find myself in times of trouble, mother mary comes to me. speaking words of wisdom, "oust Duterte"	Political Stance
@ABSCBNNews sinara mo track record mo nung naging tuta ka ni duterte	Filipino Political News

After Receiving the dataset labeled by the Domain Experts, it is once again preprocessed just as before the LDA topic modeling step. Table 6 shows the result of the final preprocessing activity.

Table 6. Definitions of stressor

Stressor	Definition
Political Stance	A side where an individual agrees based on the ideology, party, and policies
Government Policies	The government's action or intent to solve a problem/issue within the country
Election	An individual preference of candidate for the political representation of the country
Filipino Political News	Reports about the latest information revolving around politics like government policies, corruption, elections, etc.
President	Information about the President's speeches, actions, and policies.
Recollection	Origins of political events that are remembered due to their impact and relevance.

2.5 Model Testing

Before model testing, the researchers looked at the dataset to determine the number of tweets per class the dataset contains.

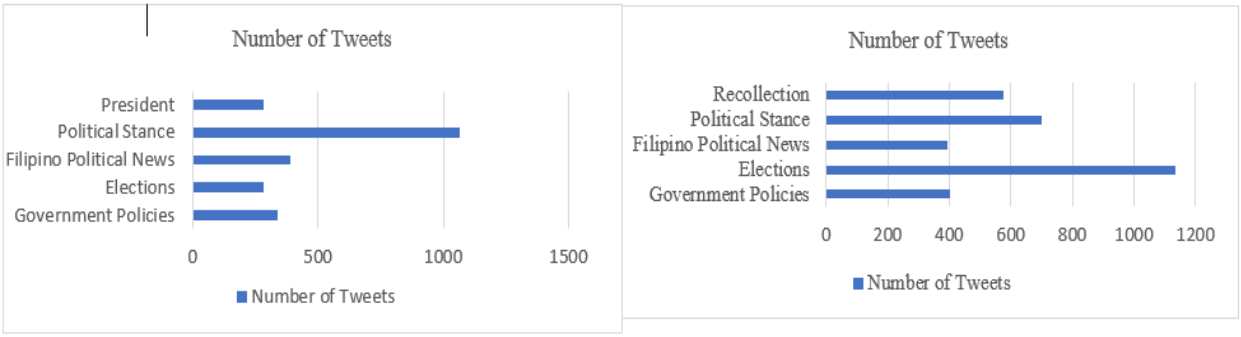

Figure 2. Numbers of Tweets for Tagalog (Left) and English (Right) Dataset

Ass shown in figure 2, the numbers of tweets per class are unbalanced so the researchers then took 280 per class on the Tagalog dataset having a total of 1,400 rows of tweets on the dataset and 390 per class on the English dataset resulting in 1,950 rows of tweets to be trained in the model. The purpose of this is to avoid biases for the model when training.

The creation of the data model will be done under Jupiter notebook on a python3 version. The model training will focus on the three algorithms of SVM, XGBoost, and Logistic Regression. The Model training will have ratios of 70/30. The researchers will run ten cross-validations of each classifier. The researchers will also use the standard statistical performance metrics like accuracy, f-score, and kappa statistics to now have better comparisons between the model but will mainly focus on the accuracy of the model.

The dataset will use the tweets as its features for the X part of the model training and the stressors for Y as it predicted target and label. The X feature will undergo a word vectorization process. The reason for this is that machine learning models don't accept string(tweets) as an input the vectorizing the words will make it into numerical values. The word vectorization method that will be used is TF-IDF, what this does is that will have Term frequency in which it will summarize how often a given word appears in a tweet and it also has an Inverse Document Frequency that downscales words that appear a lot across the tweets. After the words go thru TF-IDF vectorization they will be put on a vocabulary

where the words will have a unique integer number assigned to them. A snippet of the TF-IDF Vocabulary is shown in table 7.

Table 7. Snippet of the TF-IDF Vocabulary

Filipino	English
{'oust': 4101, 'duterte': 1450, 'kita': 2779, 'perception': 4419, 'fact': 1663, 'checker': 904, 'dilawan': 1312, 'legit': 2947, 'problem': 4659, 'might': 3498, 'addressed': 179, 'across': 165, 'industry': 2375}	'ph': 4179, 'utot': 5584, 'mo': 3261, 'nyo': 3749, 'kayo': 2458, 'lng': 2745, 'pwde': 4469, 'magsalita': 2897, 'magtanggol': 2904, 'pag': 3849, 'bayaran': 539, 'agad': 160, 'gago': 1654, 'utak': 5581}

The Y labels are consisting of the five stressors which are also strings because of this it will go to the process of Label Encoder. This is done to transform categorical data of string type in the dataset into numerical values which the models will accept. Since there are 5 stressors Y labels will have 5 numerical values each representing the stressors, the numerical values will represent as 0, 1, 2, 3, 4 which 0 will represent the first stressor then 1 will represent the second stressor then 2 will represent the third stressor then 3 will represent the fourth stressor and the 4 will represent the fifth stressor. In table 8, it shows the encoded stressor label.

Table 8. Encoded Stressor Labels

Filipino		English	
Stressors	Labels	Stressors	Labels
Political Stance	3	Elections	0
Government Policies	2	Government Policies	2
Elections	0	Recollection	3
Filipino Political news	1	Filipino Political News	1
President	4	Political Stance	4

SVM
The use of the SVM is chosen by the researchers because it is a good predictive analysis for data classification. In most cases, it's a binary classifier but the SVM the researchers use is a multiclass SVM since we have five target variables that were fitted in Y in the data modeling process for this algorithm.

XGBoost
The XGBoost is a decision tree type of machine learning algorithm, it uses a gradient boosting framework. This is a good algorithm for the researcher's dataset because decision-based tree algorithms tend to work well on small to medium size data.

Logistic Regression
Logistic Regression is one of the most well-known algorithms to be used for classification problems. Logistic Regression is used to predict a data value based on their feature's prior features. For this study, we are using multinomial Logistic Regression since we are trying to create a model that can predict five classes.

SVM and XGBoost have undergone hyperparameter tuning with the use of grid search to improve the performance of the model.

3. Result and Discussions

3.1. Summary of Result

Based on table 9, the accuracy score shows that it has acceptable results score in terms of predicting. The models show between 70% in terms of accuracy. All three Models shows identical results in terms of their Accuracy and F-Score the reason for this is that each Machine learning Algorithms do well when it comes to multi-class predictions, The reason why the accuracy might not have been higher is that that tweet is used for the X feature and there are a lot of variations of tweets within the researcher's dataset

Table 9. Performance for Annotated Tagalog with classification dataset

Annotated Dataset Tagalog		
Model	**Accuracy**	**F1 score**
SVM	70%	71%
XGB	70%	70%
Logistic Regression	71%	71%

Table 10. Performance for Annotated English with classification dataset

Annotated Dataset English		
Model	**Accuracy**	**F1 score**
SVM	69%	70%
XGB	72%	72%
Logistic Regression	71%	71%

In table 10, the result in the English dataset is somewhat di same as the Tagalog dataset in terms of their accuracy score which is around 70%. Both the English and Tagalog dataset seems to have the same output because both datasets use tweets as the X feature and a lot of these tweets has different variations of it.

3.2 Tagalog Words on NLP Methods and Machine Learning

Even with the presence of Tagalog words, NLP methods used (LDA) were able to identify the commonly used words in the Filipino language and cluster them to form topics together which was later evaluated and labeled by the Domain Experts to form the stressors. For the model, since classifiers on machine learning does not accept a string as an input, the difference of language is not seen since the text are vectorized and the ML models then look for the pattern which would then predict the stressors wherein all the models were able to predict with SVM performing the best out of the three.

3.3 English Model performance vs Tagalog Model performance

The English model did perform similarly with the Tagalog dataset mainly because both contained the language of the other. Taglish (Tagalog + English) was seen on both datasets as code-switching is prevalent in the Philippines. This could mean that trying to mine tweets of the different languages on the same topic of the area will inevitably fall to code-switching.

3.4 Error Analysis

These are some reasons why the data models only produce acceptable results and not a high one based on their performance.

Confusion Matrix: Figure 3 shows the results for the Confusion Matrix of the best model.

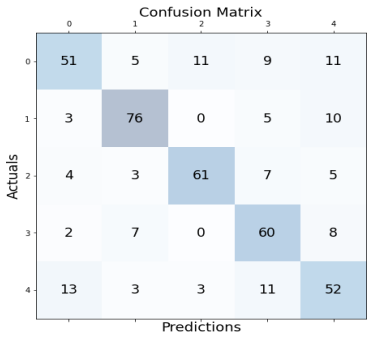

Figure 3. Confusion Matrix LogReg

The confusion matrix shows that for the 1st topic (Elections), the best model predicted 51 of the tweets to be true positives while the majority of the false negatives were labeled as part of Government Policies and President. This may be because Presidency is correlated to Elections. The 2nd topic (Filipino Political news) was predicted with 71 true positives with the majority again on the President Topic. The 3rd topic (Government Policies) is predicted with 61 true positives while it was even out on the false negatives among other topics. The 4th topic (Political Stance) was predicted with 60 true positives with the majority of the false negatives being among topics President and Political news. The last topic President yielded 52 true positives and the majority of the false negatives are with topic Elections and Political stance.

Misleading Tweets: Some tweets that were collected are considered satire this can cause for word pooling of the topic models to be inconsistent. Verbal irony such as sarcastic tweets may cause some inconsistencies in the collection of words.

Inconsistent Topic Models: The creation of topic models saw that some models have a very similar set of words within them this can cause the model to have a hard time predicting the stressor accurately.

Code-Switching: Both datasets containing the language of the other (both containing Taglish tweets) cause the most influential variable in the process. Rendering the dataset is virtually the same.
Prediction from other tweets: Predicting tweets from a range of topics outside of politics would prevent the model from predicting accurately since the stressors are only political-based.

4. CONCLUSION

This research sees that you can find stressors based on a collection of Tagalog tweets or an English tweet that has a Filipino-based hashtag. It fulfills its role in detecting the suitable stressor for a particular tweet based on the hashtag using the methods of topic modeling. It also saw have a method to create a data model for stressor detection and the prediction that would have 70-71% performance among the three models, The reason why each model is like each other is that the algorithm of the model primarily focuses on multiclass predictions. The model that performed the best was Logistic Regression (71%) in our study because the algorithm prioritizes posterior class probability and is very effective on a linear classification problem.

For future work, the researchers recommend mining tweets in different countries in the same generalized topic if possible since mining the same topic on the same area (area-specific topics) to compare two different languages will fall to code-switching. The researchers also suggest using a better cleaning method if there is one created in terms of Tagalog text because data cleaning for Tagalog text seems to be minimal compared to other languages and does not have a dedicated function. The researchers also suggest having other similarities-based methods to find stressors and other NLP techniques to make a cluster for the stressor creating. Also, try different methods such as a neural network for the data model creation.

Reference

Abastillas, G., You Are What You Tweet: A Divergence in Code-Switching Practices in Cebuano and English Speakers in Philippines, *Mehta S. (eds) Language and Literature in a Glocal World*, Springer, Singapore, 2018.

Armstrong, S., Philippine Tsismis: Gossip and the Politics of Representation in Jessica Hagedorn's Dogeaters Postcolonial Text, *Postcolonial Text*, Vol 16, No 4, 2021.

Cooper, C. L. and Quick, J. C., *The Handbook of Stress and Health: A Guide to research and Practice*, John Wiley Sons Inc., 2017.

De Leon, G. F. and Lintao, R., The Rise of Meme Culture: Internet Political Memes as Tools for Analysing Philippine Propaganda*, Journal of Critical Studies in Language and Literature*, 2(4), 1-13, 2021.

Gollapalli, S. D. and Li, X., Using PageRank for Characterizing Topic Quality in LDA, In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR '18),* Association for Computing Machinery, 2018.

Guntuku, S. G., Buffone, A., Jaidka, K., Eichstaedt, J. C., & Ungar, L. H., Understanding and Measuring Psychological Stress Using Social Media, *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01), 214-225, 2019.

Hussein, D., A survey on sentiment analysis challenges, Journal of King Saud University - Engineering Sciences, Volume 30, Issue 4, Pages 330-338, 2018.

Jose, N. Chakravarthi, B. R., Suryawanshi, S., Sherly, E. and McCrae, J. P., A Survey of Current Datasets for Code-Switching Research, *Proceedings of 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),* pp. 136-141, 2020

Negara, E.S., Triadi, D., and Andryani, R., Topic Modelling Twitter Data with Latent Dirichlet Allocation Method, *Proceeding of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS),* 386-390, 2019.

Pillai, G. R., Thelwall, M. and Orasan. C., Detection of Stress and Relaxation Magnitudes for Tweets, *Proceedings of the The Web Conference 2018 (WWW '18)*, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1677–1684, 2018.

Rijhwani, S., Sequiera, R., Choudhury, M., Bali, K., and Maddila, C. S., Estimating Code-Switching on Twitter with a Novel Generalized Word-Level Language Detection Technique, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics,* Association for Computational Linguistics, 1971–1982, 2017.

Sjöström, S., *Labelling theory in Routledge International Handbook of Critical Mental,* Taylor & Francis Group, 2017.

Seaward, B. L., *Managing stress: Principles and strategies for Health and Well Being*. Burlington, MA: Jones & Bartlett Learning, 2018.

van der Van der Schuur, W.A., Baumgartner,S.E., Sumter, S.R., Social Media Use, Social Media Stress, and Sleep: Examining Cross-Sectional and Longitudinal Relationships in Adolescents, *Health Commun*, 34(5):552-559, 2019.

Wang, X., Zhang, H., Cao, L., and Feng, L., Leverage Social Media for Personalized Stress Detection, *Proceedings of the 28th ACM International Conference on Multimedia,* Association for Computing Machinery, New York, NY, USA, 2710–2718, 2020.

Biographies

Mark Gabriel E.Edaño is a Computer Science Student at Mapua University who is currently in his final year of graduation for his course. Artificial Intelligence was his specialization of the subject under the university, and he also studied pattern recognition and technopreneur ship as his electives. He was a software/data analyst intern at the company of NCSI Philippines. He strives to be a data scientist particularly big data that are sports-related.

Ryan Joseph S. Gonzales is an undergraduate student of Mapua University that is taking a Bachelor of Science in Computer Science. He specializes in Artificial Intelligence and is currently in his final year of graduation. He has done his Internship at Chimes Consulting as a Backend Developer and is planning to pursue a career as a software engineer. His interest revolves around the automation of non-supervised functions/projects that can be applied to a larger scale task.

Raphael Carlo B. Laguda is an undergraduate student of Mapua University, taking BS Computer Science. He specializes in Application Development and also studied pattern recognition and technopreneurship as his electives. He was a Software Engineer intern in Realtair Inc. His interest is in Web and Application Development, Artificial Intelligence, and Game Development.

Joel de Goma is a student of Mapua University, taking PhD in Computer Science. He also an instructor of the said University.