Short Paper*

Evaluating Urgency of Typhoon-Related Tweets Through Sentiment Analysis Using Artificial Neural Networks

Ronnel P. Ermino
College of Computer Studies and Multimedia Arts, FEU Alabang, Philippines
erminoronnel@gmail.com
(corresponding author)

Ray Carlo A. Abacan
College of Computer Studies and Multimedia Arts, FEU Alabang, Philippines
raabacan@feualabang.edu.ph

Nadine Gweneth C. Diamante
College of Computer Studies and Multimedia Arts, FEU Alabang, Philippines
nadinegwenethdiamante@gmail.com

Kristine Marie N. Faca
College of Computer Studies and Multimedia Arts, FEU Alabang, Philippines
kristinefaca@gmail.com

Thatiana Erica D. Juntereal
College of Computer Studies and Multimedia Arts, FEU Alabang, Philippines
thatianajuntereal@gmail.com

Recommended citation:

Abstract

Purpose – The main objective of this project was to evaluate typhoon-related Tweets' urgency using sentiment analysis with supervised learning over an artificial neural network.

Method – The researchers implemented artificial neural network and natural language processing techniques for sentiment analysis and evaluation of the urgency score of typhoon-related Tweets. The model's accuracy on training and validation was evaluated simultaneously. A separate validation using 100 data was done using confusion matrix analysis.

Results – The accuracy of the model in training was at 99.87% and the loss was 0.0074. Validation was conducted simultaneously with the training. It was found that the accuracy of the model was at 99.17% and the loss was 0.0680. The confusion matrix analysis showed that the sensitivity was 97.67% and the specificity was 100%. The positive predictive value was 100% and the negative predicted value was 98.28%. Both false positive and false discovery rates are at 0% while the false-negative rate was at 2.33%. Finally, the F1 score was 98.82% and accuracy was 99%.

Conclusion – The implementation of the architecture of the model was successful; the researchers concluded that the training produced successful results by looking at the high accuracy prediction of the model and the low loss during the simultaneous training and validation, and confusion matrix analysis for the separate validation.

Recommendations – The researchers recommend expanding the vocabulary of the model by adding more diverse data to the dataset when training. The model produced by this study can be used in incident reporting systems that will be helpful during times of typhoon-related disasters.

Research Implications – Using the model produced by the study in incident reporting applications of the government and NGOs will be more efficient than manually looking at typhoon-related posts on Twitter.

Keywords – Artificial Neural Networks, Sentiment Analysis, TensorFlow, Urgency Score, Supervised Learning

INTRODUCTION

Background of the Study

According to the World Health Organization, about 100,000 people die each year as a result of natural disasters such as earthquakes, volcanic eruptions, storms, tsunamis,

flooding, wildfires, and droughts, while more than 150 million are affected (Pappas, 2020). This equates to 0.1 percent of all deaths worldwide (Ritchie & Roser, 2019). The Philippines is visited by an average of twenty typhoons every year, five of which are destructive. The year 2020 has been especially tough for the Filipino people because there was a monthly episode of calamity. November 2020, was the worst when five consecutive typhoons hit some parts of Luzon that caused floods (Palatino, 2020). Super Typhoon Rolly (Super Typhoon Goni) and Typhoon Ulysses (Typhoon Vamco) have affected 1.6 million people. To add to the devastation, some parts of the country were severely flooded and people were trapped on the roof of their households (American Red Cross, 2020).

Technology plays a vital role in the prevention, planning, and management of disasters. When it comes to emergency preparedness and mitigation, communication, data collection, and protection of human welfare are critical, and technology is a key component of these solutions. Many tools, including social media and other online platforms, are used by federal and local governments to maintain open lines of communication (Milosh, 2020). New technologies have been developed over the years to enhance the productivity and efficacy of first responders, expanding the role of technology in disaster relief (Parrell, 2016). While national emergency numbers are being used by many, the industry has lagged in adapting to the most recent technical advances. This inability to accept change undermines the efficacy of responders, emphasizing the need for emerging technology to support all people. (Onion et al., 2019)

In the Philippines, in terms of disaster preparedness, local governments can use GeoMapperPH, a GIS-based tool, to create evacuation routes, develop evacuation centers, and situate public facilities far from danger-prone areas. (Mina, 2021). Social networks are one of the most effective means for teams operating in the middle of a crisis (such as government social media teams or health care professionals) to get authoritative facts to the public quickly (Cooper, 2020). A study conducted in the city of Padang in Indonesia (Carley et al., 2016) found that Twitter was used more than other social networks in the local community context. Based on the study's findings, it was suggested that early disaster warnings on Twitter can alert people. The Eriksson and Olsson study (2016) reported that citizens view Twitter as a tool to alert users in a crisis, find current events and locate news media coverage. Hashtags make Twitter a useful tool in crisis communication because hashtags allow users to gather all the Tweets about a specific topic into one area making it easier to find breaking news and share with a wider group of readers interested in that topic (Eriksson, 2012; Eriksson and Olsson, 2016; Spence et al., 2015). This shows how social media is a powerful asset in emergency management as it allows information to be passed to the intended audience rapidly and efficiently (Sharp, 2020).

Even though social media platforms are powerful assets in emergency management, there is still a gap on what people in the Philippines are using in actuality and the standard modes of communication and protocols in the government. Right now,

the government cannot standardize social media in information forwarding during disasters (Peneyra, personal communication, 2021). But there are foundations and NGOs that utilize specialized modes of communication during emergencies, one good example is the Philippine Disaster Resilience Foundation (PDRF) (Peneyra, personal communication, 2021). Amidst all the advancements in technology, the Philippines is still having a hard time when it comes to rescuing operations. Hence, the reason for the researchers to do further studies and decided to focus on the application of sentiment analysis with artificial neural networks on social media data, specifically Tweets, to evaluate the urgency of the said Tweets to help with the rescue operations of NGOs and the government during or after typhoon-related disasters.

Purpose

The main objective of this project is to evaluate typhoon-related Tweets' urgency using sentiment analysis with supervised learning over an artificial neural network. The specific objectives are: (1) to preprocess the research's dataset for training and testing using natural language processing techniques like tokenization and converting sentences to sequences or vectors, (2) to build an artificial neural network with an embedding layer, global average pooling layer and an output layer with Sigmoid activation function for sentiment analysis to be used for the research. (3) To evaluate the performance of the artificial neural network's prediction using Binary Cross-Entropy and Confusion Matrix Analysis.

LITERATURE REVIEW

The study of Kusumasari & Prabowo (2020) titled: "Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific" focused on the different uses of Twitter during disasters in Asia and the Pacific in 2014 and 2015. The purpose of the study was to show the pattern of use of Twitter to send warnings and identify crucial needs and responses. Results cast light on not only how various types of users utilize Twitter in times of disaster but also on how several potential Twitter users are absent during disasters. Twitter use for relief coordination occurs understandably in the aftermath of a disaster, but the speed and reach of Twitter make it an ideal platform for disaster preparedness coordination and planning.

In addition to this, in the study "The power of social media during natural calamities", Filipinos used mass media to create hashtags like #RescuePH and #ReliefPH in the height of Habagat in 2012. Now, these hashtags have become a unified hashtag for disaster response. The National Disaster Risk Reduction and Management Council (NDRRMC) showed a general lack of disaster preparedness and urgency. No high alert and no massive preemptive evacuation were undertaken before "Ulysses" ("Vamco"), unlike Typhoon "Rolly" ("Goni"), as gleaned from the NDRRMC Twitter feed and situation reports. Citizens helped by reporting on-ground situations which NDRRMC used

for their Rescue mapping efforts. This was then shared with the Office of Civil Defense (Lardizabal-Dado, 2020). The contribution of this article to the study is the confirmation that crowdsourced data is useful in rescue operations

A study by Takahashi et al. (2015) titled, "Communicating on Twitter during a disaster: An analysis of Tweets during Typhoon Haiyan in the Philippines" analyzed the composition of Tweets during and after Typhoon Haiyan hit the Philippines. The Tweets gathered for the study came from local and out-of-country users. It was revealed that 85.6% of the Tweets were in English (Takahashi et al., 2015). One of the contributions of the said study was that it revealed that most Tweets during the disaster were in English, which helped researchers decide to include English Tweets for training to yield better accuracy.

In a study named, "Use of Social Media Data in Disaster Management: A Survey", social media has played a significant role in disaster management, as it enables the general public to contribute to the monitoring of disasters by reporting incidents related to disaster events. Big Data technology is a key technology for social media data management due to the high volume of generated social media data. Moreover, machine learning and information retrieval algorithms are widely used to collect, classify, and extract essential information from social media (Phengsuwan, 2021).

In a study named: "Sentiment Analysis of Typhoon Related Tweets using Standard and Bidirectional Recurrent Neural Networks", the main purpose of the study is to identify the sentiments expressed in the Tweets sent by the Filipino people before, during, and after Typhoon Yolanda using two variations of Recurrent Neural Networks; standard and bidirectional. Supervised training was done on two types of Recurrent Neural Network algorithms, standard and bidirectional Recurrent Neural Networks (Imperial, 2019). This was the closest related study to this research. Although Imperial's (2019) models focused on classifying Tweets and not evaluating them, sentiment analysis was the method that was the same method for this research. The difference with the model was that Imperial's was Recurrent, and this research used Feed Forward Artificial Neural Network. This research's focus was on evaluating the urgency score of Tweets expressed through a number between 0 and 1 while Imperial's (2019) research focused on classifying Tweets as positive, neutral, or negative.

METHODOLOGY

Project Design

The researchers started with gathering data from Twitter using Tweepy and Twitter API. The data from Twitter was labeled with 1 when urgent and 0 otherwise. The labeled dataset was then preprocessed by removing duplicates, lowercasing, removal of non-alpha characters, tokenization, and converting the data to sequences. Preprocessing

was necessary before it can be used in training and validation. The clean, preprocessed data was used for training and validation of the model with an input layer, embedding layer, global average pooling layer that averages the values from the embedding, hidden layer which consists of a dense layer with a Rectified Linear Unit activation function to avoid the vanishing gradient problem and speed up the training process, and an output layer that was implemented using TensorFlow 3 and Python. After training, the model and the vocabulary were saved.
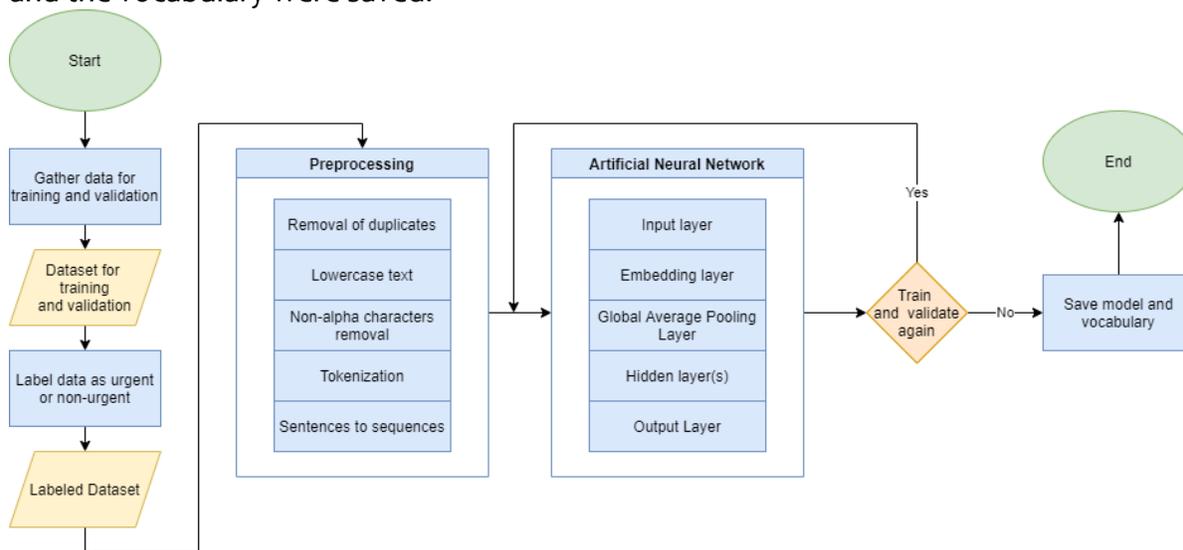

Figure 1. Process flow diagram of the project.

Methods for Evaluation of the Model

Binary Cross-Entropy evaluates how well the model classifies and evaluates the input. This loss function returns a low value when a model makes a good prediction and a high value when the model makes a bad prediction. Aside from Binary Cross-Entropy in the evaluation of the performance of the model's prediction, the Confusion Matrix was used to analyze the accuracy of the prediction. The accuracy of the model can be evaluated by adding the true positives and true negatives and finding the ratio of the sum to the total number of predictions (Rajan, 2020). Since the accuracy of the prediction was imperative to the study, Confusion Matrix analysis was used. Additionally, sensitivity, specificity, precision, negative predictive value, false-positive rate, false discovery rate, false-negative rate, and F1 score was computed.

RESULTS

Dataset

The researchers gathered the Tweets related to typhoon Ulysses from November 8 to November 16 of the year 2020 (Table 1). The Tweets' language was either English,

Tagalog, or a combination of both. For training and validation, there was a balanced number of urgent and non-urgent Tweets. The total of 4,020 Tweets consisted of 2,010 urgent and 2,010 non-urgent Tweets. 3,900 of the Tweets were used to train the model and the remaining 120 were used for validation.

Table 1. Dataset specifics

Language(s)	English, Tagalog
Origin	Tweets related to typhoon Ulysses from November 8 -16, 2020
Number of Tweets	4020
Number of urgent Tweets	2010
Number of non-urgent Tweets	2010
Number of Tweets for training	3900
Number of Tweets for validation	120

Architecture

The architecture of the Artificial Neural Network had an Input layer with an input size of 140, which was the maximum number of words that a Tweet can contain (Figure 2). The next layer was the Embedding layer with 8 dimensions and 64000 parameters which were 8 dimensions multiplied by the maximum number of the words in the vocabulary which was set to 8,000. It was followed by the Global Average Pooling layer with an output shape of 8. The hidden layer that followed consisted of 216 parameters, 24 nodes, and Rectified Linear Unit as the activation function. The output layer had an output shape of 1 and a Sigmoid activation function.
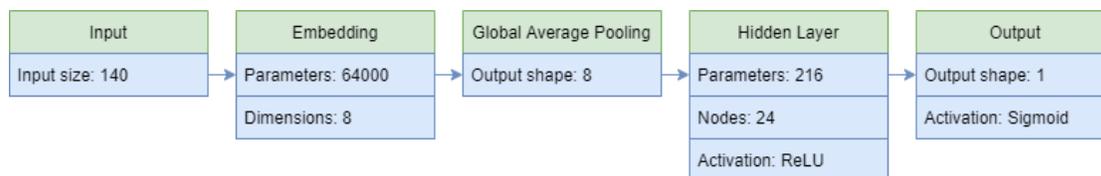

Figure 2. Model layers.

Training and Validation

Table 2 shows two Disaster-Related Tweets with the evaluated urgency score using the trained model. The first Tweet produces high urgency score and is leaning to 1 because it is asking for rescue and needs urgent help. The second Tweet produces a low urgency score and is leaning to 0 because it is not urgent since it is about donation drives

only. The model's prediction accuracy during training reached 99.87% with a loss of 0.0074. The validation accuracy reached 99.17% and a loss of 0.0680 (Table 3, Table 4, and Table 5).

Table 2. *Disaster-Related* Tweet with Urgency Score

Tweet	Urgency Score
Please help us nalulunod na po kami at ang taas na po ng baha	0.9998237
We have donation drives to help our kababayans! Send us a message if you want to donate!	0.0000019

Table 3. Accuracy and Loss

Training accuracy	Training loss	Validation accuracy	Validation loss
99.87%	0.0074	99.17%	0.0680

Table 4. Confusion Matrix

n = 100	Prediction: Urgent	Prediction: Non-Urgent
Actual: Urgent	42	1
Actual: Non-Urgent	0	57

For another separate validation, the researchers used 100 Tweets with 43 actual urgent and 57 actual non-urgent divisions. True positives were 42 and no false-positive; true-negative were 57 and 1 false-negative (Table 4).

Table 5. Confusion Matrix analysis

Sensitivity	97.67%
Specificity	100%
Positive Predictive Value (Precision)	100%
Negative Predictive Value	98.28%
False Positive Rate	0%
False Discovery Rate	0%
False Negative Rate	2.33%
F1 Score	98.82%
Accuracy	99%

After the confusion matrix was analyzed, the sensitivity was identified to be 97.67% and the specificity to be 100% (Table 5). The positive predictive value was 100% and the negative predicted value was 98.28%. Both false positive and false discovery rates are

at 0% while the false-negative rate was at 2.33%. Finally, the F1 score was 98.82% and accuracy was 99%. The model was saved in three formats, Protocol Buffers or Protobuf (.pb), Hierarchical Data Format or HDF5 (.h5), and JSON (.json) for web applications using JavaScript and (TensorFlow.js).

DISCUSSION

According to Carol Smith (personal communication, June 19, 2021), a senior research scientist in human-machine interaction at Carnegie Mellon University, the target accuracy for predictions concerning emergency response was about 90% to 95%, it is because of the nature of the application of the model which was incident reporting to be followed by an emergency response. The model was able to breach the target range with a 99.87% accuracy on training and 99.17%. By looking at the difference between the accuracy of training and validation, the researchers interpreted that there was no significant evidence to say that there was overfitting or underfitting. The loss of the training which was at 0.0074 was an indication that the training was successful and that the difference between predicted and actual urgency scores were little. The loss of 0.0680 in the validation was also an indication that there was little difference in the predicted and actual urgency scores on the data that the model was not trained. Both losses were under the threshold of 0.693, so it was an overall success for the training and validation. The confusion matrix analysis further solidifies the success of the training and validation. With only one false prediction over a hundred predictions, the accuracy was at 99%. Naturally, other forms of analysis of the confusion matrix produced favorable results. The sensitivity was at 97.67%, specificity and precision were at 100%, the negative predictive value was at 98.28%, false positive and false discovery rate was at 0%, false-negative rate was at 2.33%, and lastly, the F1 score was at 98.82%. Although with a different context, this research's model accuracy which is 99.17% is greater than Imperial's (2019) 86.99% and 87.6% accuracy for standard and bidirectional RNN.

CONCLUSIONS AND RECOMMENDATIONS

The researchers were successful in producing a preprocessed dataset for the training of the model. The implementation of the architecture of the model was also successful, the researchers concluded that the training produced successful results by looking at the high accuracy prediction of the model and the low loss during training and validation and confusion matrix analysis. Even though the researchers were able to reach a high accuracy using the model and the preprocessed dataset but it is still recommended for future researchers to expand the vocabulary of the embedding layer by adding more diverse data to the dataset, even more so if the model will process multilingual inputs. The model produced by the study is also recommended to be integrated into applications for incident reporting. The researchers recommend the use of JavaScript and Node JS in live streaming Tweets and predicting the urgency score using the model through TensorFlow.js with the JSON model format. Applications using the model can also be implemented in Python through the HDF5 and Protobuf model format.

IMPLICATIONS

If the model will be further trained with more diverse data in the future, the model's vocabulary can be expanded, and predicting urgency scores of new data will improve. Additionally, if the model will be used in an application in incident reporting by the government and NGOs, finding calls for help on Twitter will be much faster and easier.

REFERENCES

American Red Cross. (2020). *Philippines: Red Cross Helps After Typhoons Goni, Vamco Strike.* Retrieved from RedCross: https://www.redcross.org/about-us/news-and-event/news/2020/red-cross-response-to-typhoon-goni.html

Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., & Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *Safety Science, 90,* 48-61.

Cooper, P. (2020). *How to Use Social Media for Crisis Communications and Emergency Management.* Retrieved from HootSuite: https://blog.hootsuite.com/social-media-crisis-communication/

Eriksson, M., & Olsson, E.-K. (2016). Facebook and twitter in crisis communication: A comparative study of crisis communication professionals and citizens. *Journal of Contingencies and Crisis Management, 24(4),* 198-208.

Imperial, J., Orosco, J., Mazo, S., & Maceda L. (2019). *Sentiment analysis of typhoon related tweets using standard and bidirectional recurrent neural networks.* Retrieved from: Cornell University: https://arxiv.org/abs/1908.01765

Kusumasari, B., & Prabowo, N. P. A. (2020). Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific. *Natural Hazards, 103*(3), 3415-3435.

Lardizabal-Dado, N. (2020). *The power of social media during natural calamity*. Retrieved from Manila Times: https://www.manilatimes.net/2020/11/22/business/sunday-business-i-t/the-power-of-social-media-during-natural-calamities/799521/

Mina, R. (2021). *Philippines looks to improve disaster preparedness with geospatial tech.* Retrieved from: L News Mongabay: https://news.mongabay.com/2021/03/philippines-looks-to-improve-disaster-preparedness-with-geospatial-tech/

Milosh, A. (2020). 10 *Roles of Information Technology in Emergency Preparedness and Prevention.* Retrieved from Atera: https://www.atera.com/blog/10-roles-of-information-technology-in-emergency-preparedness-and-prevention/#:~:text=Technology%20plays%20a%20key%20role,prime%20part%20of%20these%20solutions.

Onion, A., Sullivan, M., & Mullen, M. (2019). *First 9-1-1 call is placed in the United States.* Retrieved from History.com: https://www.history.com/this-day-in-history/first-911-emergency-call-is-placed-in-the-united-states

Palatino, M. (2020). *A deadly and disastrous 2020 for the Philippines*. Retrieved from The Diplomat: https://thediplomat.com/2020/12/a-deadly-and-disastrous-2020-for-the-philippines/

Pappas, S. (2020). *Top 10 deadliest natural disasters in history*. Retrieved from Live Science: http://livescience.com/33316-Top-10-deadliest-natural-disasters.html

Parrell, W. (2016). *In search and rescue, tech can make the difference between life and death*. Retrieved from Digital Trends: https://www.digitaltrends.com/outdoors/how-search-and-rescue-find-lost-people-in-the-wilderness/

Phengsuwan, J., Shah, T., Thekkummal, N. B., Wen, Z., Sun, R., Pullarkatt, D., ... & Ranjan, R. (2021). Use of social media data in disaster management: A Survey. *Future Internet*, *13*(2), 46. Retrieved from MDPI: https://www.mdpi.com/1999-5903/13/2/46

Rajan, S. (2020). *Analyzing the performance of the classification models in machine learning*. Retrieved from Toward Data Science: https://towardsdatascience.com/analyzing-the-performance-of-the-classification-models-in-machine-learning-ad8fb962e857

Ritchie, H., & Roser, M. (2014). *Natural disasters. Our world in data.* Retrieved from Our World Data: https://ourworldindata.org/natural-disasters.

Sharp, E. N., & Carter, H. (2020). Examination of how social media can inform the management of volunteers during a flood disaster. *Journal of Flood Risk Management*, *13*(4), e12665.
Spence, P. R., Lachlan, K. A., & Rainear, A. M. (2016). Social media and crisis research: Data collection and directions. *Computers in Human Behavior*, *54*, 667-672.

Takahashi, B., Tandoc Jr, E. C., & Carmichael, C. (2015). Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior*, *50*, 392-398.