ASIAN ASSOCIATION for CRISISONOMY

Proceedings of 2022 Asian Conference on Crisisonomy
(20 ~ 23 July, 2022; Jeju National University; Asian Association for Crisisonomy)

Natural Disasters: Machine Learning and Tweet Classification

Anish Joshi[1+], Heriz Bista[2#], Naman Subedi[3*], Ishwar KC[4], Gajendra Sharma[5]

[1]Department of Computer Science and Engineering, Kathmandu University, Dhulikhel, Nepal

ABSTRACT

Social-media has been the go-to source of news for the modern masses in recent years. The flow of information is unmatched on these sites where big multinationals maintain servers to handle the user information. Twitter has proved a great tool for its use in the spatial and temporal modeling of events which can be very efficient in gaining automated details on Natural Disasters. Using the Machine Learning technique for Natural Language Processing (NLP) and Twitter, all of this information can be leveraged to create a kind of natural disaster identifier. This research paper primarily focuses on the use of public tweets to assess the occurrence and impact of natural disasters. At first, the tweets from the public accounts on Twitter were extracted and filtered to disaster-type tweets using various disaster keywords. Then Bidirectional Encoder Representation from Transformer (BERT) model was used to classify tweets to check if the tweets were actually of the disaster.

Key words:NLP, Disaster, BERT, Twitter

1. Introduction

The volume of people using social media is increasing every year. With 4.62 billion (Chaffey, 2022) people using social media which is more than half of the entire population of the world (58.4%) (Chaffey, 2022) the power of social media in information and data collection becomes unprecedented. Popular social media sites like Twitter which has about 217 million monthly active users (4th Quarter 2021) (Statista, 2022). Around 500 million tweets are sent each day (Sayce, 2020) on Twitter which makes Twitter a very important platform to gather information. Twitter can prove to be a very useful tool to extract information and collect an extensive amount of data in different cases like crisis or disaster management (Dufty, n.d.). The participation of normal people through social media like Twitter in the aftermath of any disaster improves the reliability of the information. (Gao, Barbier, & Goolsby, 2011)

With such a large amount of crowd-sourced data running through a single stream and an API to access it, Twitter has become synonymous with unlabeled datasets for all kinds of deep learning applications. This is especially true for Natural Language Processing; hence we have used Twitter as a source on which to base our findings. Through the use of Tweepy, a Twitter API extension, we were able to extract tweets containing keywords relating to natural disasters. Hence the process of data sourcing was made effortless through the use of Twitter and its over 300 million users (Statista, 2022) take on subject matter from every aspect of life.

The collective knowledge of the whole twitter base is an assemblage of humor, news, attempts at marketing, and misinformation campaigns. To help us get through this noise and conduct analysis on real actionable intel, we make use of the BERT model to classify the tweets collected. Although being time-consuming to train and infer plus having over 300 times the parameters of CNN, Bert is better for use in NLP (Zhu, 2021) than most CNNs in text classification tasks. Using web scraping with the information acquired from Twitter, it is possible to get a spatial (location, latitude, and longitude) overview plus the temporal (time of the day, date of the event) rundown of the disaster in question. Hence this purposed system that we have built aims to give technical insight into the possibilities of crowdsourcing data with the power of deep learning.

2. Related Work

Research on natural disasters before and after the fact is an age-old interest due to its grave nature and the possibility to save lives. With the rise of the internet, humanity has a new faucet for research on natural disasters and more chances to make an impact.

Many a resource has been developed from disaster datasets (Disaster Tweets, 2020) and (Unknown, 2021). On top of that, Kaggle, a data science website hosts other resources for deep learning enthusiasts like competitions, datasets, code, and discussion forums. Competitions like (Unknown, n.d.) and (Unknown, n.d.) help collect solutions and experiments from different individuals in the same vicinity. The datasets provided help researchers train their model without having to go through the tedious task of hand labeling data while competitions help sharpen and compare their methods with others in the business.

Prior to this research paper, similar research had been conducted (Sit, Koylu, & Demir, 2019) which employed Long Short-Term Memory (LSTM) networks for text classification instead of the Bidirectional Encoder Representation from Transformers (BERT) model as we did. Although this may be because BERT wasn't fully-fledged and released by the time this paper had been published. While (Sit, Koylu, & Demir, 2019) made a specific case study on Hurricane Irma, our research is a more holistic approach to natural disasters, their frequency, and sensing using deep learning.

The traditional approaches to such undertaking would be Logistic Regression (Bewick, Cheek, & Ball, 2005) Support-vector machines (Unknown, n.d.)l and naïve Bayes (NB) classifiers (Rish, 2001). The LSTM approach as mentioned in (Sit, Koylu, & Demir, 2019) is a quite favored technique, slightly out phased only by the introduction of the BERT, RoBERTa and alike models.

Also as shown in (Kongthon, Haruechaiyasak, Pailai, & Kongyoung, 2014) the previous research papers on related topics focused more on a single disaster instance like the 2011 Thai Flood. This type of study can be very helpful to recognize factors affecting the ramifications of certain types of disasters and how they are communicated by the larger general public. The specific adoption of Twitter for these kinds of studies had been ever-increasing as shown in (Hughes & Palen, 2009) and (Sakaki, Okazaki, & Matuso, 2010).

Although the number of studies performed and papers written on this subject topic encompassing both the validity of Twitter as a substantial news source and BERT's ability to excel in inference tasks are plentiful, there are still some key questions that need answering. One of the main ones we have aimed to answer is "Can Twitter perform as a disaster sensing source with the help of deep learning?".

Solutions to text classification provided in the aforementioned competitions like (C., 2021) and (X., 2020) provide helpful insight into different implementations. Deep learning has no correct answer, it's like cooking as an art, there will always be a younger prodigy who is better at it than you. Tinkering is an integral part of deep learning as so many before us have shown in (C., 2021), (X., 2020) , and (Sit, Koylu, & Demir, 2019) with different approaches to preprocessing, model training, and so on.

3. Methodology

In this experiment, Tweepy, a python library, was used for listening to tweets containing disaster-related keywords. Keywords used were [Earthquake, Storm, Tsunami, Cyclone, Rain Storm, Rain, Hurricane, Flood, Tornado] The tweets fetched consisted of many sub-properties of which the time of the tweet created and the actual tweet text data was saved. The entire process was divided into 3 Phases.

Phase I: The first phase of the experiment is to get live tweets using the Twitter API. They are acquired using our disaster keywords. This data is then stored in the database for further preprocessing and inference.

Phase II: The second phase of the project is the most important part. Here tweets data from the database is taken and preprocessed and passed through the BERT model to get an inference of whether it is real or fake.

Phase III: The third phase consists of calculating outliers from the peaks on the data.

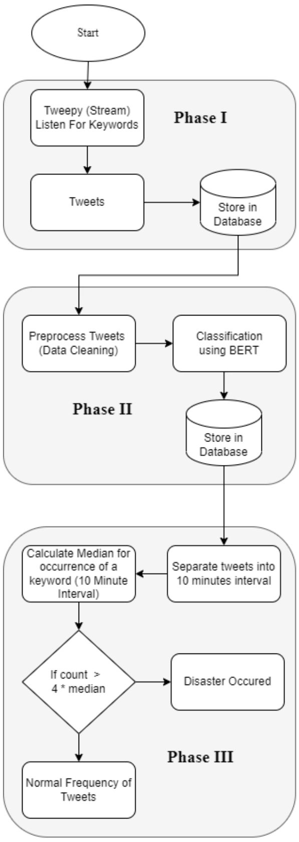
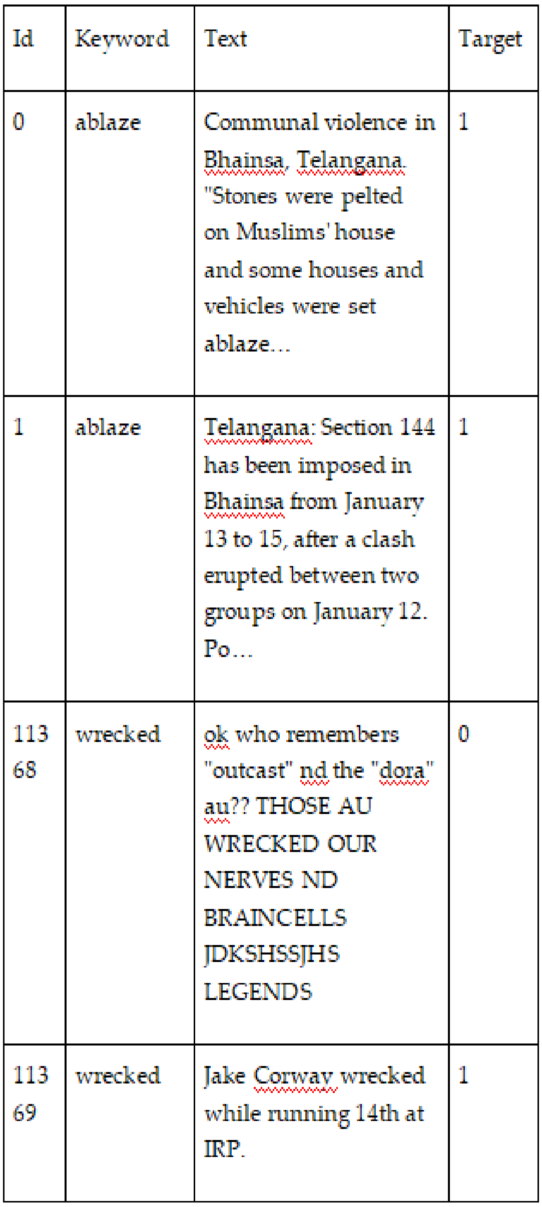

Start
Tweepy (Stream) Listen For Keywords
Phase I
Tweets
Store in Database
Preprocess Tweets (Data Cleaning)
Classification using BERT
Phase II
Store in Database
Calculate Median for occurrence of a keyword (10 Minute Interval)
Separate tweets into 10 minutes interval
If count > 4 * median
Disaster Occured
Phase III
Normal Frequency of Tweets

Figure 1: Phase Diagram

2.1 Training Dataset

Kaggle is the most often used platform for downloading datasets. The labeled dataset utilized to train the model was obtained from Kaggle named Disaster Tweets (Disaster Tweets, 2020). The dataset contains 11,000 tweets associated with natural and artificial disasters, 219 of them. The target field contains the integer values 0 or 1 where 0 signifies the tweet is not about the concerned disaster while 1 denoted the tweet is about the disaster in question. Only 2114 of the total 11,000 tweets are labeled as 1 whereas the left is annotated as 0. The Id field contains a unique identifier for each tweet to help section them off as a test set and training set before training the model (Table 1).

Table 1: Test set and training set



Id	Keyword	Text	Target
0	ablaze	Communal violence in Bhainsa, Telangana. "Stones were pelted on Muslims' house and some houses and vehicles were set ablaze…	1
1	ablaze	Telangana: Section 144 has been imposed in Bhainsa from January 13 to 15, after a clash erupted between two groups on January 12. Po…	1
11368	wrecked	ok who remembers "outcast" nd the "dora" au?? THOSE AU WRECKED OUR NERVES ND BRAINCELLS JDKSHSSJHS LEGENDS	0
11369	wrecked	Jake Corway wrecked while running 14th at IRP.	1

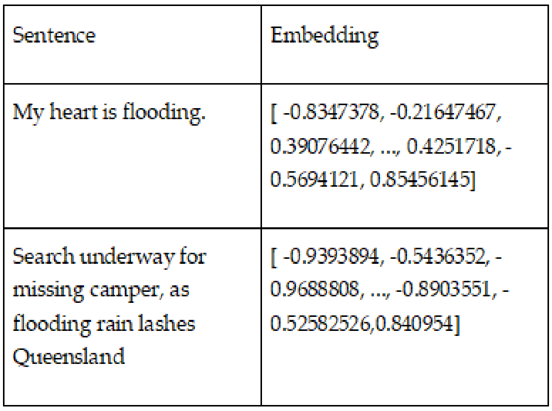
2.2 Preprocess data

Fetched tweets consisted of different tags punctuation marks, URLs, typos, spacing, retweet properties, and many others. So before training machine learning models on natural language texts in tweets, tweets were preprocessed to remove stop words and word tokenization.

2.3 Bert for Binary Classification

Fetched tweets were in the form of natural language. And for machines to understand it Natural Language Processing (NLP) was required. This is where the BERT comes in. It is an open-source machine learning framework that is based on a transformer architecture. A transformer is a deep learning model in which every output element is connected to every input element, and the meaning between them is dynamically calculated based on their connection (Table 2).

Table 2: Different contextualize embeddings.



Sentence	Embedding
My heart is flooding.	[-0.8347378, -0.21647467, 0.39076442, …, 0.4251718, -0.5694121, 0.85456145]
Search underway for missing camper, as flooding rain lashes Queensland	[-0.9393894, -0.5436352, -0.9688808, …, -0.8903551, -0.52582526,0.840954]

Bert can efficiently generate contextualized embedding vectors. Here the word 'flooding' is common but the meaning for both the sentences are totally different hence the values of embedding for 'flooding' in both the sentences will be different based on their context.

2.3.1 Bert Model Architecture

Bertbase which consists of 12 encoders and 768 Forward Feed Network and 12 attention heads is used for this experiment.

CLS (Classification) is the special token and first token of every sequence. The hidden state corresponding to this token is used as the mean sequence representation for classification. In other words, CLS represents the meaning of the entire sentence.

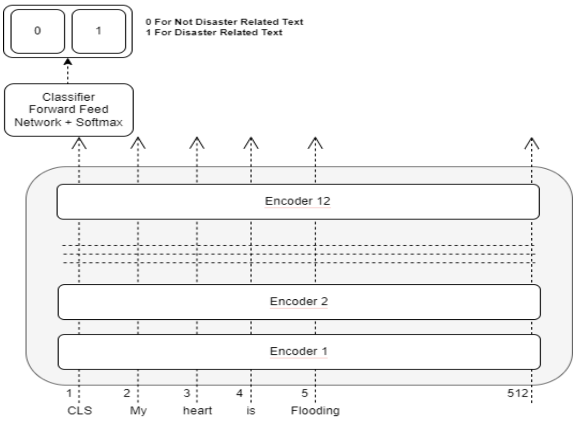

0
1
0 For Not Disaster Related Text
1 For Disaster Related Text
Classifier Forward Feed Network + Softmax
Encoder 12
Encoder 2
Encoder 1
1 CLS
2 My
3 heart
4 is
5 Flooding
512

Figure 2: BERT architecture for disaster classification

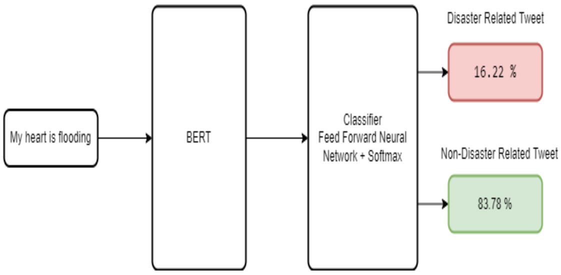

My heart is flooding
BERT
Classifier Feed Forward Neural Network + Softmax
Disaster Related Tweet
16.22 %
Non-Disaster Related Tweet
83.78 %

Figure 3: Block Diagram of BERT for disaster classification

2.4 Analyzing the processed data

The tweets after being classified with the help of the BERT model were then divided into the 10 minutes interval according to the time-stamp of the tweets. The total number of tweets containing the specific keyword like “earthquake”, “hurricane”, “flood”, etc. in that 10-minute interval were counted and stored according to the keywords in the database along with the date and time-stamp. and the median was calculated using the formula below.

$$Median = L + \frac{(\frac{n}{2} - c.f.)}{f} \times i$$

where,

n = number of items in the set

L = lower limit of median class, median class is the class where (n/2)th item is lying.

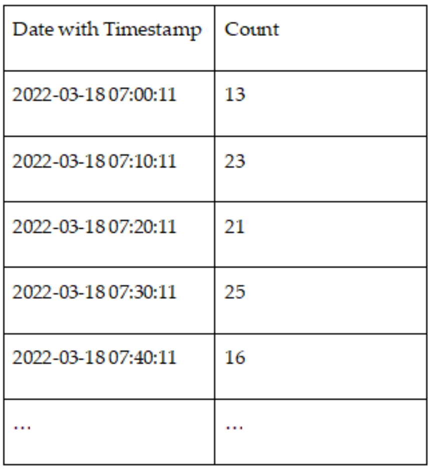
c.f. = Cumulative frequency of the class preceding the median class.

f = Frequency of median class

i = Class interval of median class.

After the median is calculated, the median of the count is also stored in the database of the particular disaster keyword. Below is the table of the Earthquake counts which is stored in the database (Table 3).

Table 3: interval of real data

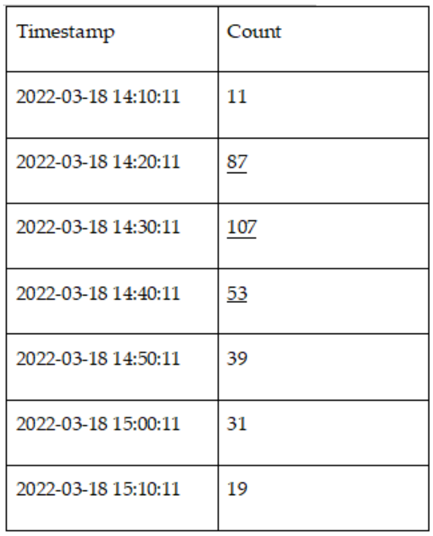


Date with Timestamp	Count
2022-03-18 07:00:11	13
2022-03-18 07:10:11	23
2022-03-18 07:20:11	21
2022-03-18 07:30:11	25
2022-03-18 07:40:11	16
…	…

Median of count for the entire data of keyword earthquake was calculated (12.0).

3.5 Information Extraction

Table 4: Abnormal points in collected data



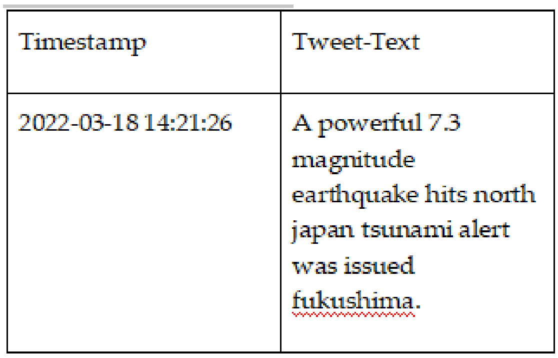
Timestamp	Count
2022-03-18 14:10:11	11
2022-03-18 14:20:11	87
2022-03-18 14:30:11	107
2022-03-18 14:40:11	53
2022-03-18 14:50:11	39
2022-03-18 15:00:11	31
2022-03-18 15:10:11	19

Table 4 shows the abnormal points in collected data.

If the count of a specific keyword in a 10-minute interval is greater than the four times the tweet count of that category (underlined), the tweets from that interval are taken and location is determined using Geotext (Unknown, 2018), a python library that distinguishes the city and the country from a text, and latitude and longitude are determined from that location using GeoPy (Unknown), a python library that determines the co-ordinates from an address, city or a country.

After calculating the location from each tweet of that particular interval, the frequently occurring location becomes the impact area where the disaster has occurred. And those relevant tweets serve as a piece of news as shown in the table below.

Table 5: Disater related tweets



Timestamp	Tweet-Text
2022-03-18 14:21:26	A powerful 7.3 magnitude earthquake hits north japan tsunami alert was issued fukushima.

The above table (Table 5) shows a disaster related tweet from the point of abnormal count.

The extracted place is Japan and from GeoPy calculated coordinates are (36.2048°,138.2529°)

4. Results

4.1 Confusion Matrix

A confusion matrix (Kulkarni, Chong, & Batarseh, 2020) is a table that is used to define the performance of a classification algorithm. The confusion matrix (Kulkarni, Chong, & Batarseh, 2020) visualizes and summarizes the performance of a classification algorithm. The confusion matrix (Kulkarni, Chong, & Batarseh, 2020) for tweet classification is shown below.

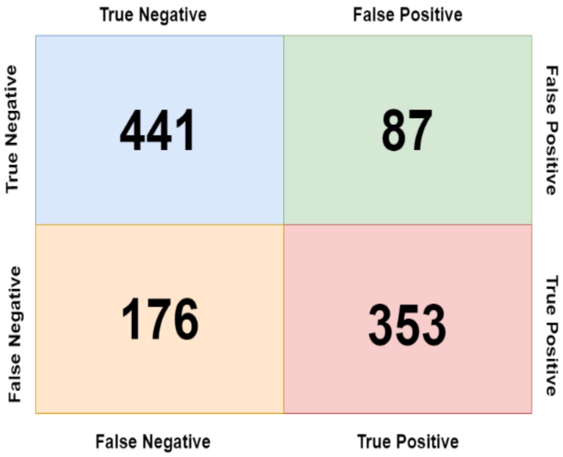


Figure 4: Confusion matrix for tweet classification

True Positive indicates the number of positive examples classified accurately. True Negative indicates the number of Negative examples classified accurately. False Negative is the number of actual Negative examples classified as Positive. False Negative is the number of actual positive examples classified as negative.

Accuracy of the model is the ratio of total correct prediction by a model and total predictions made by the model. The accuracy of a model (through a confusion matrix) is calculated using the given formula below.

$$\text{Accuracy} = \frac{TN+TP}{TN+FP+FN+TP}$$

TP stands for True Positive
TN stands for True Negative
FP stands for False Positive
FN stands for False Negative

Accuracy = 0.7511825922 * 100 = 75.11 %

4.2 Classification Report

The classification report visualizer (Unknown, n.d.) talks about the precision, the recall, the F1 score, and the support scores for the model.

Precision is the ratio of the correct classification made by the model to the total positive made by the model (either True Positive or False Positive). Precision actually determines how definitive a model is for classifying a selected trial piece positively.

$$\text{Precision (P)} = \frac{TP}{TP+FP}$$

TP stands for True Positive
FP stands for False Positive
Precision = 353 / (353 + 87)
= 0.8022727 * 100 = 80.22 %

Recall is the ratio between the actual positive samples that are correctly classified as positive to the total number of positive samples in the dataset. Recall can also be referred to as the measure of the classifier's completeness.

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP+FN}}$$

Mathematically, recall is defined as

TP stands for True Positive

FN stands for False Negative

Recall = 353 / (353 + 176)

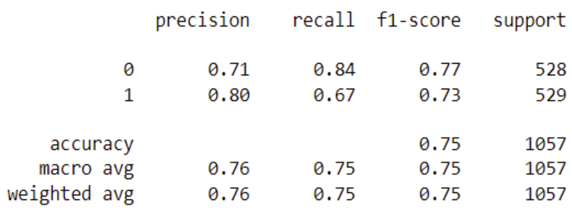
=0.667296 * 100

= 66.73 % ≈ 67 %

The F1 score is a called harmonic mean of precision and recalls such that the best score is 1.0 and the worst is 0.0.

$$\text{F1 score (F1)} = \frac{2 \times (\text{P} \times \text{R})}{(\text{P+R})}$$

P stands for precision
R stands for recall
F1 = (2 *80.22*66.33) / (80.22+66.73)
= 72.8 % ≈ 73 %



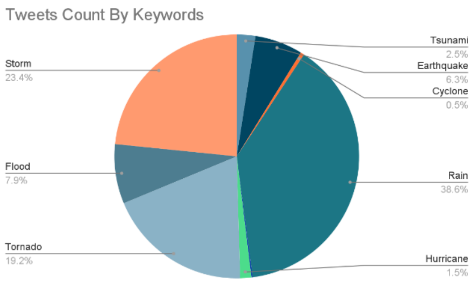
```
              precision    recall  f1-score   support

           0       0.71      0.84      0.77       528
           1       0.80      0.67      0.73       529

    accuracy                           0.75      1057
   macro avg       0.76      0.75      0.75      1057
weighted avg       0.76      0.75      0.75      1057
```

Figure 5: Table showing classification report of the model

4.3 Total Keywords

Total of 9 keywords were used and a total of one million seven hundred thousand tweets were collected between March 18, 2022 to March 27, 2022

The above pie chart was implemented which shows the frequency of tweet counts by keywords which were fetched from Mar 18, 2022 to Mar 27, 2022.



4.4 Graphs

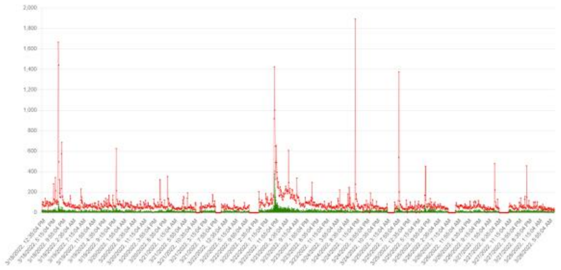


Figure 7: Graph showing processed and unprocessed tweets

The above graph shows the frequency of processed tweets and unprocessed tweets.

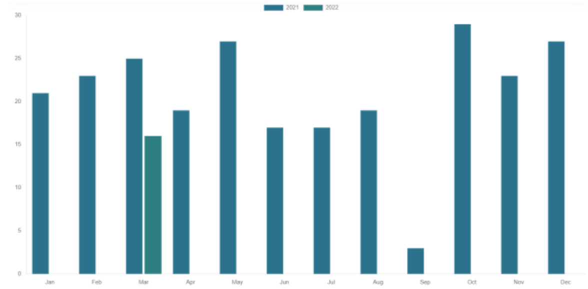


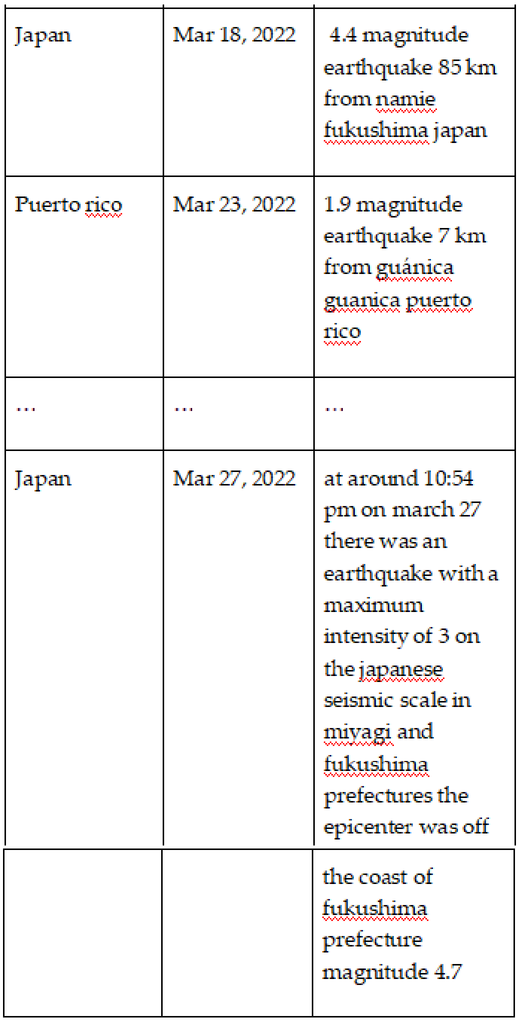
Figure 8: Bar graph showing comparison with the previous year

The above bar diagram shows the comparison of data with the previous year (2021) on March (18-27).

4.5 Tweets as News

Table 6: sample of news

Place	Date	Tweet
Japan	Mar 18, 2022	Magnitude 7.4 earthquake kills at least two people and leaves thousands without power in fukushima japan



Japan	Mar 18, 2022	4.4 magnitude earthquake 85 km from namie fukushima japan
Puerto rico	Mar 23, 2022	1.9 magnitude earthquake 7 km from guánica guanica puerto rico
…	…	…
Japan	Mar 27, 2022	at around 10:54 pm on march 27 there was an earthquake with a maximum intensity of 3 on the japanese seismic scale in miyagi and fukushima prefectures the epicenter was off the coast of fukushima prefecture magnitude 4.7

Table 6 shows the sample of news extracted from tweets.

5. Conclusion

In this paper, we have explored our pipeline to solve the Natural Language Processing task by using traditional machine learning models as well as a pre-trained universal language model. By intensively conducting experiments using BERT, we have demonstrated that BERT and our fine-tuning strategy are highly effective in text classification tasks. This has shown the efficacy of BERT and similar models in specific use cases like disaster sensing.

References

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. Critical Care, 9(1), 112. doi:10.1186/cc3045

C. (2021, September 28). Disaster tweets. From Kaggle: https://www.kaggle.com/code/crismolav/disaster-tweets

Chaffey, D. (2022). Global social media statistics research summary 2022. From Smart Insights: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/#:~:text=More%20than%20half%20of%20the,social%20media%20is%202h%2027m.

Disaster Tweets. (2020, November 12). From Kaggle: https://www.kaggle.com/datasets/vstepanenko/disaster-tweets

Dufty, N. (n.d.). AJEM Apr 2016 - Twitter turns ten: its use to date in disaster management | Australian Disaster Resilience Knowledge Hub. From Knowledge Hub: https://knowledge.aidr.org.au/resources/ajem-apr-2016-twitter-turns-ten-its-use-to-date-in-disaster-management/.

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. IEEE Intelligent Systems, 26(3), 10-14. doi:10.1109/mis.2011.52

Hughes, A., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management, 6(3/4), 248-260. doi:10.1504/ijem.2009.031564

Huiji Gao, G. B. (n.d.). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. IEEE Xplore.

Kongthon, A., Haruechaiyasak, C., Pailai, J., & Kongyoung, S. (2014). The Role of Social Media During a Natural Disaster: A Case Study of the 2011 Thai Flood. International Journal of Innovation and Technology Management, 11(03), 1440012. doi:10.1142/s0219877014400124

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). 5-Foundations of data imbalance and solutions for a data democracy. Data Democracy, 83-106. doi:https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Work Empir Methods Artif Intell.

Sakaki, T., Okazaki, M., & Matuso, Y. (2010). Earthquake shakes Twitter users. Proceedings of the 19th international conference on World wide web - WWW '10, 851-860.

Sayce, D. (2020, December 16). The Number of tweets per day in 2020. From DSayce: https://www.dsayce.com/social-media/tweets-day/

Sit, M. A., Koylu, C., & Demir, I. (2019). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. International Journal of Digital Earth, 12(11), 1205-1229. doi:10.1080/17538947.2018.1563219

Statista. (2022, May 5). Twitter: number of monetizable daily active users worldwide 2017-2021. From Statista: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/

Unknown. (n.d.). 1.4. Support Vector Machines. From scikit-learn: https://scikit-learn.org/stable/modules/svm.html

Unknown. (2018, July 30). geotext. From PyPI: https://pypi.org/project/geotext/

Unknown. (2021, September 11). NLP with Disaster Tweets - cleaning data. From Kaggle: https://www.kaggle.com/datasets/vbmokin/nlp-with-disaster-tweets-cleaning-data

Unknown. (n.d.). Classification Report — Yellowbrick v1.4 documentation. From Scikit: https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html

Unknown. (n.d.). Natural Language Processing with Disaster Tweets | Kaggle. From Kaggle: https://www.kaggle.com/competitions/nlp-getting-started

Unknown. (n.d.). Titanic - Machine Learning from Disaster | Kaggle. From Kaggle: https://www.kaggle.com/competitions/titanic

Unknown. (n.d.). Welcome to GeoPy's documentation! — GeoPy 2.2.0 documentation. From Geopy.

X. (2020, 11 07). Disaster NLP: Keras BERT using TFHub. From Kaggle: https://www.kaggle.com/code/xhlulu/disaster-nlp-keras-bert-using-tfhub/notebook

Zhu, J. (2021, December 31). Text Classification: How BERT boost the performance. From Medium: https://medium.com/walmartglobaltech/text-classification-how-bert-boost-the-performance-e65d1d678afb

Profile

anish.joshi098@gmail.com,herizbista@gmail.com,naman.subedi12@gmail.com, ishwarkc133@gmail.com, gajendra.sharma@ku.edu.np