ENG 4000: Final Gate 6

Project: Disaster Tweets - Real or Not: Natural Language Processing

Date: April 12th, 2021

Team P

Binte Zehra	215624141
Neena Govindhan	212137600
Jessie Leung	215985948
Jonas Laya	214095715
Paul Sison	214447510

Table of Contents

List of Figures

List of Tables

Executive Summary

The purpose of this document is to present the final product for the ENG 4000 final Capstone submission. The final product is **Rumble**, a website aggregating crowdsourced disaster related Tweets and labelling them as about a real disaster or not. **Key deliverables** of the project include:

1. A machine learning model using a Natural Language Processing pipeline to predict whether a Tweet is about a real disaster or not ("Real or Not? NLP with Disaster Tweets" Kaggle challenge)
2. A website to present reliable aggregate disaster information extracted from Tweets in a manner that is consumable by the general public
3. Filters for user interaction such that users can parameterize input metrics to control aggregate disaster information displayed to them on the website

The dissemination of fake news on the Internet is currently hugely problematic in that it misleads the public in a far-reaching and lasting manner. In the modern era, social media platforms such as Twitter generate an immense amount of crowdsourced data. The system novelty of Rumble is that it retrieves this vast amount of data and classifies it such that aggregate information from these Tweets can be usefully consumed by the general public. The usefulness of the website is that there is a wealth of free and accessible disaster related information available on Twitter, but a user cannot extract true and valid data just by consuming individual Tweets.

As such, the project seeks to automate truthful news extraction from social media posts in a timely manner. The innovation of Rumble lies primarily in its **confidence metric** that is generated by Machine Learning classification. This metric denotes the plausibility of Tweets being about a real disaster, and so seeks to act as an indicator label about both real disaster news and fake disaster news.

A technical risk of the project is the potential mismatch between the technical sophistication of the model and stakeholder trust. To address this challenge, Rumble provides easy-to-read metrics and filters for users to parameterize the disaster related data. The implementation and design of the product seeks to complement technical implementation with a presentation of the product and its results in a manner that is easily consumable by stakeholders. This also lends itself to the possibility of Rumble having potential useful applications beyond original use, since its design is to allow users to parameterize important inputs such as a time frame, confidence level, engagement data. The modifiability of these filters allows users of Rumble to design their own use cases for the aggregate data presented.

Furthermore, possible evolutions of the system include optimizing the model and system so that in times of emergency, actionable disaster information can be reliably accessed by first responders. The importance of this task is that in occurrences of natural disasters, immediate situational information can aid in the highly time-sensitive rescue operations of first responders. In disaster situations, even a marginal advantage could result in one or multiple lives being saved, or the prevention of catastrophic damages.

Introduction

In times of emergency, Twitter and social media have become an important communication channel for people to stay updated on what is going on in the world. However, extracting useful information from Twitter can be difficult due to the noisiness of social media platforms. The crowd-sourced nature of Twitter means that it provides instant, first-hand eyewitness accounts of live events, but also that it can be cluttered by spam, misinformation, and fake news.

The purpose of this project is to explore the usefulness of Twitter as a source of information. Ultimately, the project objective is to combine openly accessible crowd-intelligence with Machine Learning such that useful and actionable information can be extracted from Tweets. To limit the scope of the project such that it can be accomplished within the eight month timeframe of the ENG 4000 capstone, the project was focused on extracting reliable information about disasters by building a Machine Learning model that uses NLP techniques. Then displaying this information via a web application in a more easy and accessible format using visualization tools like maps and using filters to search through the tweets. This allows the users to use the information as an aid in disaster or to relay the information to those in need with reliable confidence.

This document first outlines the Technical overview, the Management volume of the project, and the Lessons Learned and Reflections. The need and scope of the project is presented in these sections. The intended audience is suggested to read this document in the following manner:

The Head of the Company/Senior Manager- The *Executive Summary* section

The Executive Summary provides a general overview of the entire project

The Customers - The *Technical Volume* section and the *Appendix A* and *B*

The Technical Volume consists of user requirements (key stakeholders, user stories and use case diagram), system requirements (product backlog), design (subsystem components and features), performance of the as-built system.

The Chief Financial Officer: The *Management Volume* section

The Management Volume consists of project management tools, Sprints, agile process review, final budget, project schedule, preliminary business case, social impact and SDGs.

Company’s continuous improvement process - The *Lessons Learned and Reflection* section

The Lessons Learned and Reflection section consists of the deviations from the plan, the failure report and lessons learned in each phase of the project.

Technical Volume

User Requirements

Presented below are the key stakeholders, the user requirements in the form of user stories and a use case diagram, illustrating the interaction between the user and the system (i.e. website).

Key Stakeholders

Stakeholder: General Public
A key deliverable of the project is a website to present reliable natural disaster information extracted from Tweets in a manner that is consumable by the general public. The best performing prediction model produced in the Alpha stage is the BERT Classifier model producing a Precision score of 0.869. While this score can still be improved in the future by tuning the model, the high precision score allows for a lower number of false positive tweets to pass through the model. This will fulfill the general needs of the current stakeholder as the model does well at not labeling tweets that are not related to disasters as true disasters. The deployed website scans the live Tweets for up-to-date disaster information, runs it through the classification model, and aggregates these labels to output reliable disaster related tweets in a manner that the general public can consume.

Stakeholder Refinement for Future Evolutions: First Responders
In the extension of the initial website, the project seeks to target first responders as a key stakeholder. Because the nature of first responder activities is such that they are highly time-sensitive and important, a higher F1-score will be necessary to justify actionable information (as this stakeholder will have a lower acceptable threshold for false positive or false negative disaster information). For first responders, even marginal situational information can aid their rescue operations such that one or many lives can be saved, or catastrophic damages prevented. However, a negative unintended consequence can be that incorrect information can lead to wastage of precious resources and rescue operation time. As such, the current classification models need to be further improved if considered again in the future.

A challenge for the future, if given a chance again, will be documenting and conveying the process of building the classification model such that first responders can trust the results of the model. To convey the criticality of this project, the methods and results of the model in gathering important natural disaster information will be presented to the stakeholder, who is a non-expert in Machine Learning, such that they are encouraged to trust the technology and adopt it as actionable information.

User Stories

ID	What does the user want?	Why does the user want this?	Functionality Implemented
US1	As a user, I want to see tweets that are talking about a disaster or not.	The user wants to receive information that is accurately describing a disaster so that accurate information can be found fast.	Implement a model that labels whether or not a tweet is referring to a disaster.
US2	As a user, I want access to real-time disaster related information that I can trust, so that I can be more informed.	The user wants to receive all the disaster related tweets of specified confidence level by querying a specific keyword.	Search Page (keyword search, display related Tweets based on engagement (likes, quotes, retweets) and confidence)
US3	As a user, I want to know if there are disaster events near me so I can get prepared and avoid getting into bad situations.	The user wants to search for an event by entering a specific location.	Add a geographic search functionality to the website
US4	As a user, I want to be able to provide help to those affected by disaster events. As a user, I might need mental health support when consuming disaster related news.	The user should have actionable access to mental health and charitable resources related to disaster events shown on our website.	Add mental health and disaster relief resources for users- GoFundMe links, hashtags for help line)
US5	As a user, I want to be provided with accurate and current/latest news	The user should see a live feed of disaster related tweets to get the latest information	Retrieve live Tweets related to disaster from Twitter and add to the model
US6	As a user, I want to see the visual representation of the tweets related to the events so I can easily evaluate their severity.	The user may want to see a visual representation of tweets distribution.	Additional event page data visualization tools (Tweet volume graphs)
US7	As a user, I want to see only Tweets that are relevant to me.	The user should be able to filter Tweets based on their personal interest (filter by confidence level, verified users, engagement, time)	Add filters to the live feed page: Confidence: 0-100 Verified: Yes or No Time Frame: Choose an option Engagement: Enter a value - Confidence filter

			implemented where users can specify a minimum confidence value on the Live Feed page, such that the feed will display only Tweets that have a confidence value higher than the specified threshold. - A filter has been added where users can specify an engagement threshold (number of Retweets, Likes and Replies) such that only Tweets with greater engagement are displayed. - Time frame filter allows the user to see Tweets within a specific time duration of their choice (i.e. Last 5 days ago) - Verified filter is implemented for users to limit the number of Tweets that shows up to only verified people and organizations (makes a Tweet more reliable)
US8	As a user, I want to view Tweets with certain keywords so that I can find Tweets related to my personal interest easily.	The user should be able to search and view Tweets according to the specified keywords.	Search for keywords has been implemented on top of filters making it even easier for users to view desired Tweets.

Table 1: User Stories

Use Case Diagram

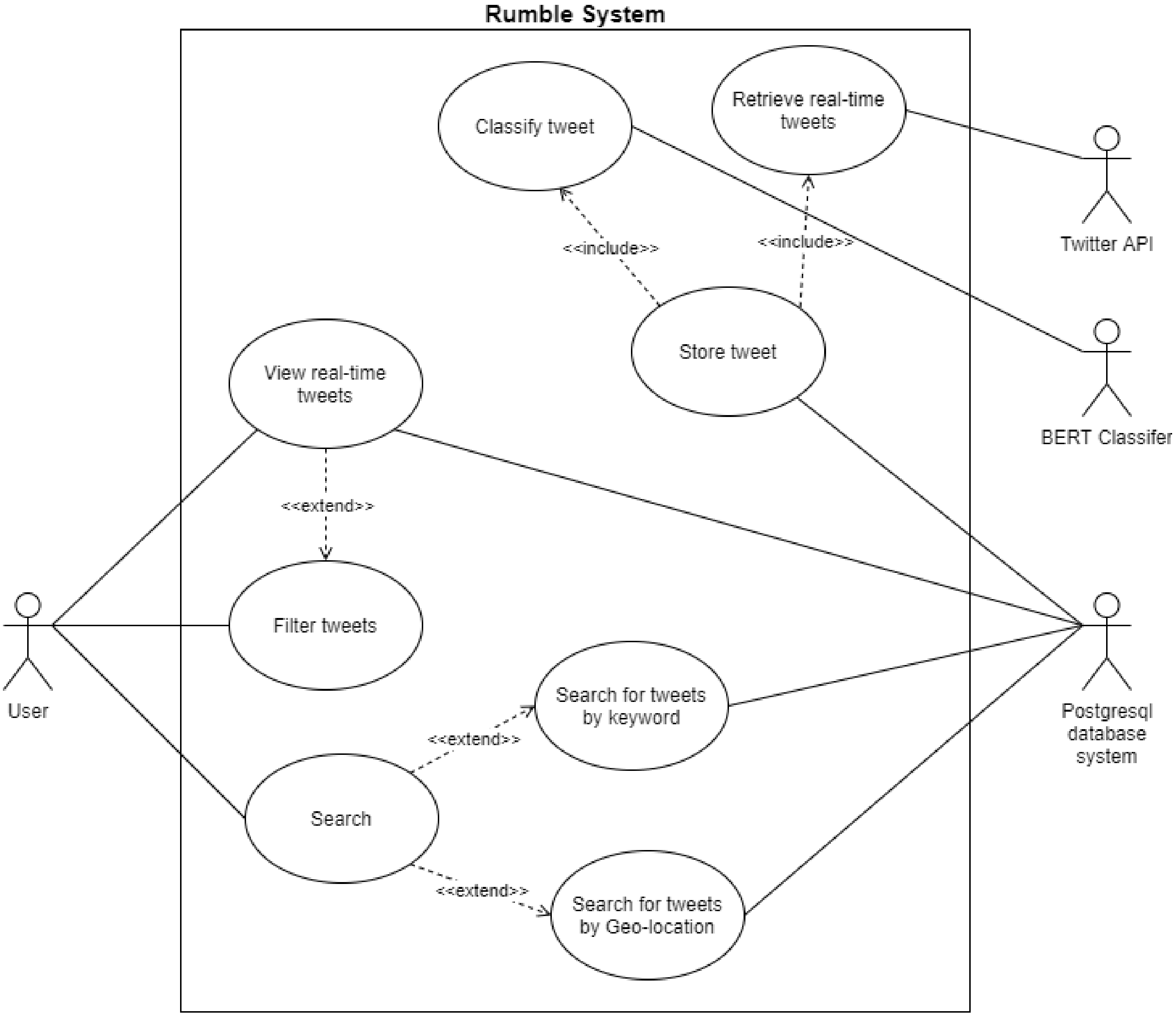


Figure 1: Use Case Diagram for the User

The use case diagram presents a high-level overview of the system under development. The diagram shows which system behaviours are relevant for each actor. For the user of the system, the relevant use cases mainly involve viewing and searching for Tweets that have been curated and/or stored by the system. The rest of the use cases represent system behaviours running in the background that are essential for delivering content and processed data to the user.

The user can search for Tweets by location on the Dashboard Page. The Dashboard Page shows a map of disaster Tweets pinned to a location. On this page the user can hover over the dots on the map to get Tweets from that location along with the percentage confidence that the tweet is related to a disaster. The geographic search functionality allows the user to filter the Live Feed of disaster related Tweets by location. The options are presented to the user in a drop-down menu (e.g. in the figure below, “Philippines” is selected). The map then zooms into the selected geographic location, and Tweets are filtered such that only Tweets from the relevant location are displayed to the user. The use case of this functionality is such that users may be primarily interested in disaster from a specified location (for example, their current city/country or their home city/country). As such, the capacity is provided for users to view disaster related Tweets within a geographic range that are of greater personal interest to the user. This page also uses data visualization tools to display the number of tweets that are disaster related and not for the past week and show the tweet distribution by confidence for the past week.

The user can use the search page to search for keywords and see the most recent of Tweets. It displays the disaster related Tweets in a visually appealing manner, and also shows the colour coded confidence level of it being a Tweet related to disaster is indicated by the percentage value, as generated by the BERT machine learning model developed in previous releases. Different filters are added to the page such that users may adjust the parameters of which Tweets they would like to see on the Live Feed page. They can filter Tweets based on the disaster related confidence level. The user can select Tweets that are only from verified accounts. Users can find Tweets within a certain time frame by choosing a duration of ‘n’ number of days. Lastly, the engagement level, which is the sum of the replies, retweets, and likes of the Tweet can also be filtered. These filters give users the added functionality of modifying the types of disaster related Tweets they would like to view on the website.

The user will be able to see Disaster Relief and Mental Health Resources section which would include links to disaster relief resources like links to nonprofit disaster relief organizations and GoFundMe pages and links to mental health resources like helplines and links to mental health organizations. In case the user gets overwhelmed by the Tweets (related to disasters), such additional resources would ensure a safe consumption of the information provided on the website.

System Requirements

Product Backlog

The product backlog consists of a **requirements review**. Each requirement is labelled with a priority level and a colour coded status indicating tracking product backlog item (PBI) completion throughout the timeline of the project. The product backlog items were added in the sequence as the project progressed and evolved. The PBIs that were not completed and to be done, if given a chance again in the future, are highlighted in grey.

The priority measure used for this backlog is as follows:

High = crucial task
Medium = important task
Low = desirable task

Legend

🟩 Completed PBIs for MVP

🟩 Completed PBIs for Alpha

🟩 Completed PBIs for Beta

🟩 Completed PBIs for Final

◼ TBC PBIs for Future

PBI#	Product Backlog	Priority
1	Introductory tutorials and familiarization with the Kaggle challenge	High
2	Begin research about natural disasters in relation to Tweets	Medium
3	Research about existing projects related to our project	Low
4	Familiarize with common ML classifiers and metrics	High
5	Familiarize with NLP approaches and techniques	High
6	Familiarize with Tweet format and metadata	Medium
7	Familiarize with Kaggle test dataset	Medium
8	Implement first model to generate predictions on Kaggle dataset	High
9	Review performance of first model	High

10	Implement and combine hand-crafted preprocessing features with classification models	High
11	Implement an LSTM prediction model	High
12	Compare model performance	High
13	Survey current stakeholders to extract social media and news content sentiment	High
14	Discuss social need and scope of project	High
15	Find additional datasets on natural disaster Tweets	Low
16	Clean and label new datasets	Low
17	Test the new dataset on model	Low
18	Combine classification models (e.g. combine non-neural models with LSTM models) to improve the model	High
19	Learn about web development and familiarize with deployment techniques	Medium
20	Implement data extraction feature to group Tweet keywords into events using Topic Modelling	Low
21	Build the website	High
22	Deploy the ML model on the website	High
23	Retrieve live Tweets related to natural disaster from Twitter and add to the model	High
24	Survey current stakeholders to get feedback on the website	High
25	Improve the model/website based on stakeholders feedback	High
26	Analyze the social impact of the project	High
27	Expand stakeholders and learn about their needs (i.e. First Responders)	Low
28	Design methods to further improve the classification models based on the data	High
29	Develop tools to allow semi-autonomous maintenance of the website	Low
30	Analyze SDG and the social impact of this project	High
31	Create a website Mock-up	High
32	Live feed table (Shows real and fake Tweets, make it visually appealing)	Medium
33	Group classified Tweets in to events based on keywords	Low
34	Event pages (word clouds (relevant information), related Tweets, total number of Tweets, probability of event of each individual tweet)	Low

35	Add a geographic search functionality to the website	High
36	Survey current stakeholders to get feedback on functionality and visual appearance	Medium
37	Deploy the website on a cloud server	High
38	Search Page (Static result table with a refresh button, keyword search, display related Tweets based on engagement (likes, quotes, retweets) and confidence of >90%)	High
39	Additional event page data visualization tools (Tweet volume graphs)	Low
40	Survey stakeholders (final product) for feedback and improve final product	High
41	Add mental health and disaster relief resources for users- GoFundMe links, hashtags for help line	Medium
42	Final fixes on the website	Medium

Table 2: Product Backlog

Design

A high level system diagram of the **as-built system design** can be found in Figure 2 below. The website (implemented using Dash framework for the front and back-end) and the PostgreSQL database storing disaster related Tweets were deployed to a cloud server using Heroku. These subsystems are deployed such that users can interact with the website via an Internet browser. The users interact with the website front-end, which communicates with the back-end to retrieve and process Tweet data from the PostgreSQL database via SQLalchemy, and then returns the called data to the user. The PostgreSQL database stores Tweets, retrieved in real-time by the Twitter API via Tweepy. The stored Tweets are then classified by the BERT model. The map is built using the Openstreetmap API which uses Geocoder to access the Nominatim API.

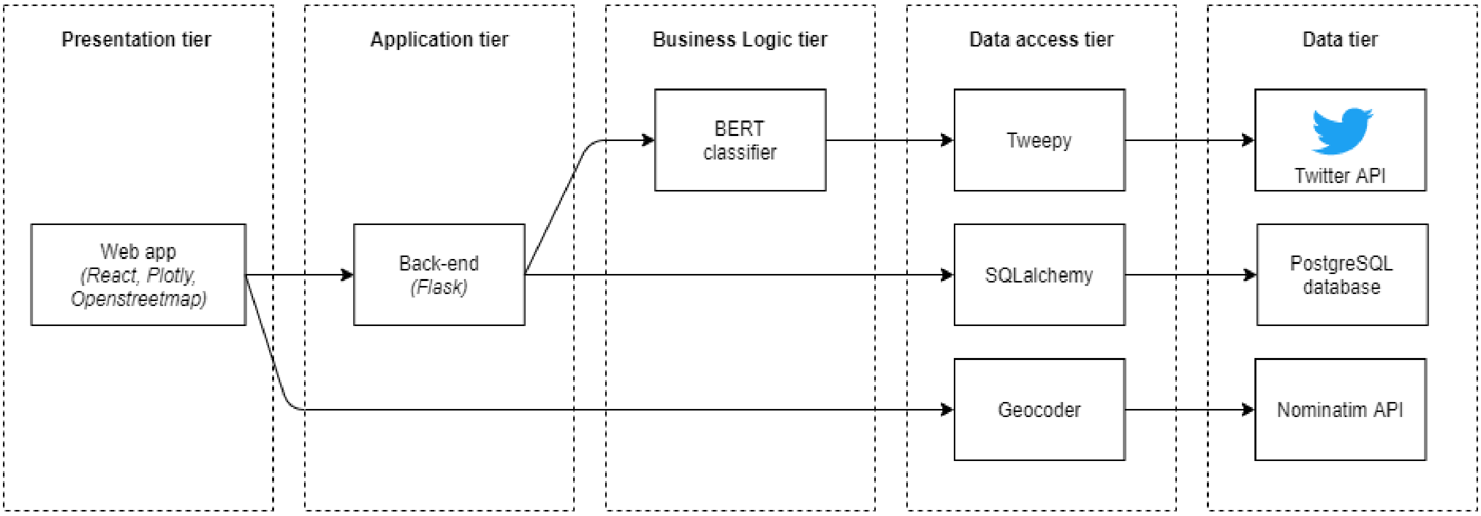


Figure 2: 5-Tier System Architecture

Subsystem Components and Features

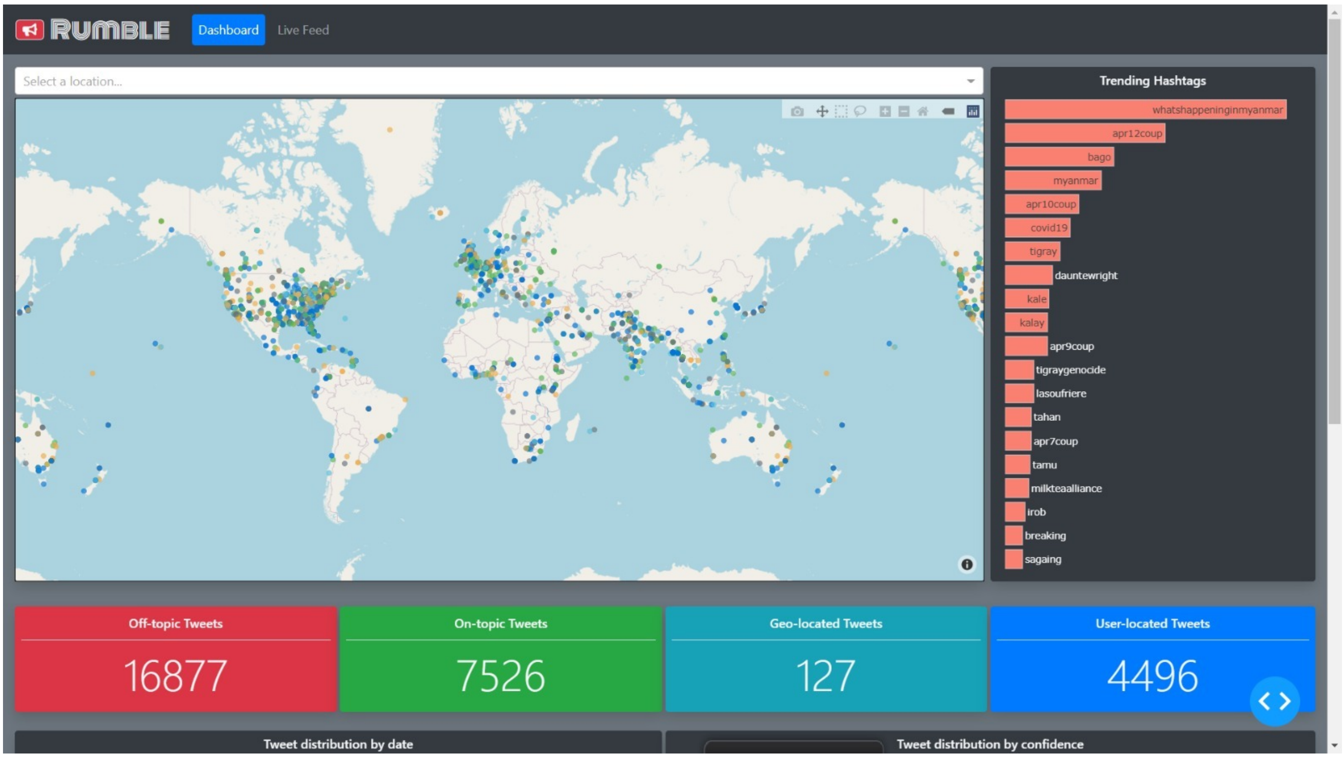
The as-built system, named **Rumble**, consists of two main pages: the Dashboard page and the Live Feed Page. The main purpose of the web application is to provide users access to aggregated information about Tweets that contain disaster related keywords (e.g. “disaster”, “volcanic eruption”, “typhoon”). A full list of disaster related keywords used to stream Tweets can be found in Appendix B.

As the main stakeholder, users typically only view individual Tweets or a feed of most recent Tweets (within a hashtag on the Twitter explore page). In the current era of “fake news” dissemination, this can be problematic for users in that they have little contextual information to base their judgement of whether a Tweet’s information is valid or false.

As a platform, Twitter content is crowdsourced through the Tweets of many individuals, yet individual consumption of Tweets can leave users vulnerable to fake news or misinformation. Rumble seeks to address this problem by aggregating disaster related information from Tweets and presenting it to users for useful and reliable consumption. Rumble presents this information using a **Confidence** metric, which is a confidence level (as a percentage) for whether a Tweet containing a disaster keyword is “On-Topic” in that it actually refers to a disaster. This confidence metric is produced by the implemented **BERT classification model**, which is further discussed in following sections of this document. The main components and features of Rumble and their contribution towards this goal are as follows.

Dashboard Page

The main Dashboard page consists of a map, a table of trending hashtags, Tweet label counters, and two Tweet distribution visualization graphs.



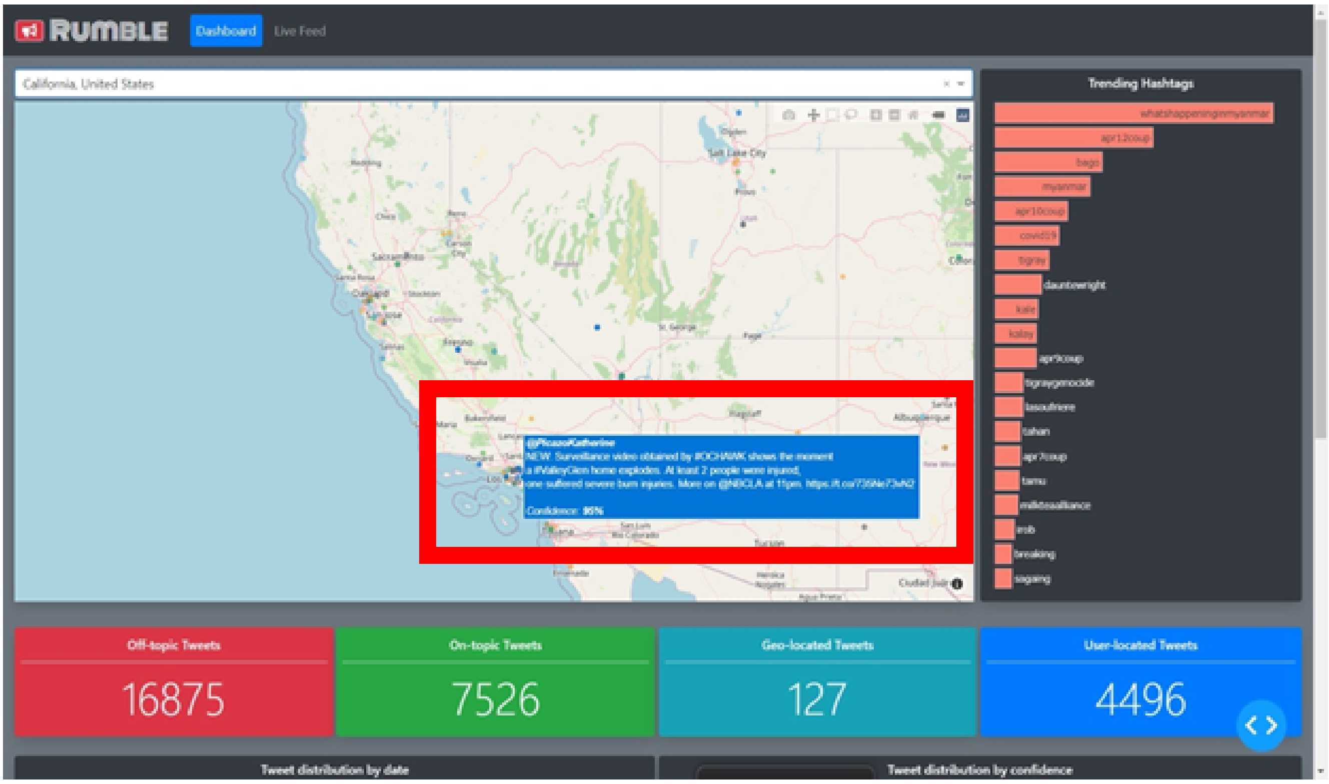
Off-topic Tweets 16877

On-topic Tweets 7526

Geo-located Tweets 127

User-located Tweets 4496

Figure 3: Rumble Main Dashboard Page



@PicazoKatherine
NEW: Surveillance video obtained by #OCHAWK shows the moment
a #ValleyGlen home explodes. At least 2 people were injured,
one suffered severe burn injuries. More on @NBCLA at 11pm. https://t.co/73SNe73vN2

Confidence: 85%

Figure 4: Rumble Map Hover Tweet Display Functionality

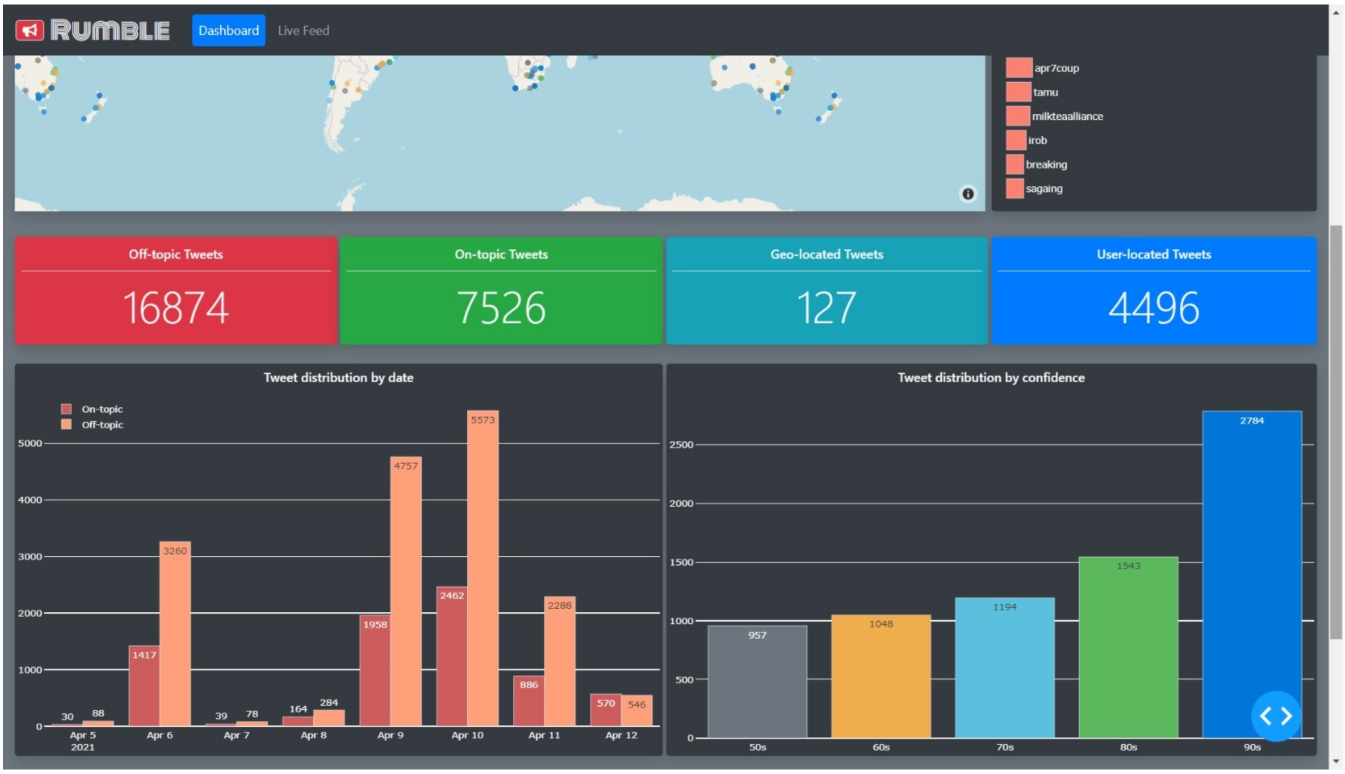


Figure 5: Rumble Tweet Distribution Visualization Graphs

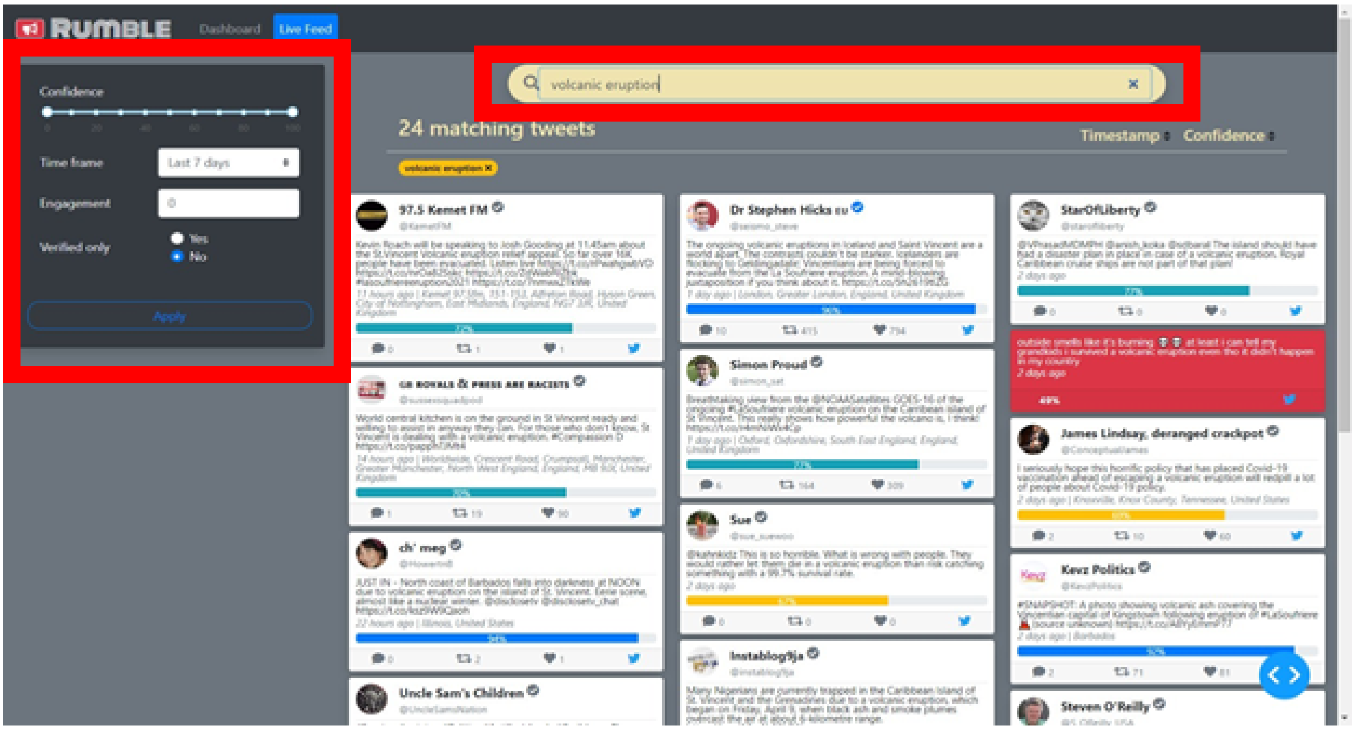
As in Figure 3 and 4 above, the first component is the Tweet Map where a location can be selected from the dropdown menu. The Map displays an indicator point where a Tweet contains a disaster related Tweet, and when the mouse is hovered over the indicator point a tooltip will popup, displaying the Tweet text as well as information such as the author's Twitter handle and Rumble's Confidence measure of the Tweet.
Next, the Trending Hashtags table displays the top trending hashtags within the aggregated Tweets that contain disaster keywords. The hashtags are filtered by a confidence level threshold (e.g. if 80% is set, the table will calculate the top hashtags based only on collected Tweets that have a minimum confidence level of 80% and disregard the other Tweets).

The counter boxes provide high-level descriptive statistics, including a comparison of the number of Off-Top to On-Topic Tweets. This comparison is helpful in showing the ratio of Tweets that are truly disaster related to those that are not, even though they contain disaster related keywords.

Similarly, the visualization graphs, in Figure 5, provide a visual to aid users visualize "Tweet Distribution by Date" and "Tweet Distribution by Confidence". These tools are useful to provide a visual timeline of aggregated disaster related Tweets and their distributions. Important patterns can be extrapolated from these visual tools. For example, if a significant global disaster occurs, the likely spike in aggregate Tweet data around that date could be visualized on the graph, and information such as whether the ratio between On and Off-Topic Tweets changes during a global disaster could be discovered.

The usefulness of these features serve two main use cases. First, users will likely be interested in disaster events that are of geographical proximity to themselves or in a geographic location that they are interested in (e.g. immigrants from Hong Kong will likely be interested in keeping up with disaster events that happen there). Therefore, the map function allows users to aggregate disaster Tweet information from a specific geographic location. Furthermore, users will also likely be interested in prominent events, which is served by the trending hashtags table. The table provides a list of "top events" that the user may or may not have prior knowledge of, thereby helping inform the user of valid prominent disaster events.

Live Feed Page



Confidence: 0 – 100
Time frame: Last 7 days
Engagement: 0
Verified only: ○ Yes ● No
Apply

volcanic eruption

24 matching tweets

Timestamp Confidence

volcanic eruption ×

97.5 Kemet FM @KemetFM
Kevin Roach will be speaking to Josh Gooding at 11.45am about the St.Vincent volcanic eruption relief appeal. So far over 16K people have been evacuated. Listen live https://t.co/nPwahgabVO https://t.co/nrCa8J5akc https://t.co/ZdWabRjTha #lasoufriereeruption2021 https://t.co/1nmeaZTkWe
11 hours ago | Kemet 97.5fm, 151-153, Alfreton Road, Hyson Green, City of Nottingham, East Midlands, England, NG7 3JR, United Kingdom
72% · 0 · 1 · 1

GB ROYALS & PRESS ARE RACISTS @sussexsquadpod
World central kitchen is on the ground in St Vincent ready and willing to assist in anyway they can. For those who don't know, St Vincent is dealing with a volcanic eruption. #Compassion D https://t.co/pappATJM4
14 hours ago | Worldwide, Crescent Road, Crumpsall, Manchester, Greater Manchester, North West England, England, M8 9JX, United Kingdom
70% · 1 · 19 · 90

ch' meg @Howlerin8
JUST IN - North coast of Barbados falls into darkness at NOON due to volcanic eruption on the island of St. Vincent. Eerie scene, almost like a nuclear winter. @disclosetv @disclosetv_chat https://t.co/ksz9W9Qaoh
22 hours ago | Illinois, United States
94% · 0 · 2 · 1

Uncle Sam's Children @UncleSamsNation

Dr Stephen Hicks @seismo_steve
The ongoing volcanic eruptions in Iceland and Saint Vincent are a world apart. The contrasts couldn't be starker. Icelanders are flocking to Geldingadalir. Vincentians are being forced to evacuate from the La Soufriere eruption. A mind-blowing juxtaposition if you think about it. https://t.co/5N2619I5ZG
1 day ago | London, Greater London, England, United Kingdom
96% · 10 · 415 · 794

Simon Proud @simon_sat
Breathtaking view from the @NOAASatellites GOES-16 of the ongoing #LaSoufriere volcanic eruption on the Caribbean island of St Vincent. This really shows how powerful the volcano is, I think! https://t.co/HmNWW4Cg
1 day ago | Oxford, Oxfordshire, South East England, England, United Kingdom
77% · 6 · 164 · 309

Sue @sue_suewoo
@kahnkidz This is so horrible. What is wrong with people. They would rather let them die in a volcanic eruption than risk catching something with a 99.7% survival rate.
2 days ago
67% · 0 · 0 · 0

Instablog9ja @instablog9ja
Many Nigerians are currently trapped in the Caribbean Island of St. Vincent and the Grenadines due to a volcanic eruption, which began on Friday, April 9, when black ash and smoke plumes overcast the air at about 6-kilometre range.

StarOfLiberty @starofliberty
@VPrasadMDMPH @anish_koka @sdbaral The island should have had a disaster plan in place in case of a volcanic eruption. Royal Caribbean cruise ships are not part of that plan!
2 days ago
77% · 0 · 0 · 0

outside smells like it's burning 💀💀 at least i can tell my grandkids i survived a volcanic eruption even tho it didn't happen in my country
2 days ago
49%

James Lindsay, deranged crackpot @ConceptualJames
I seriously hope this horrific policy that has placed Covid-19 vaccination ahead of escaping a volcanic eruption will redpill a lot of people about Covid-19 policy.
2 days ago | Knoxville, Knox County, Tennessee, United States
69% · 2 · 10 · 60

Kevz Politics @KevzPolitics
#SNAPSHOT: A photo showing volcanic ash covering the Vincentian capital of Kingstown following eruption of #LaSoufriere ⚠ (source unknown) https://t.co/ABYy0mmF77
2 days ago | Barbados
92% · 2 · 71 · 81

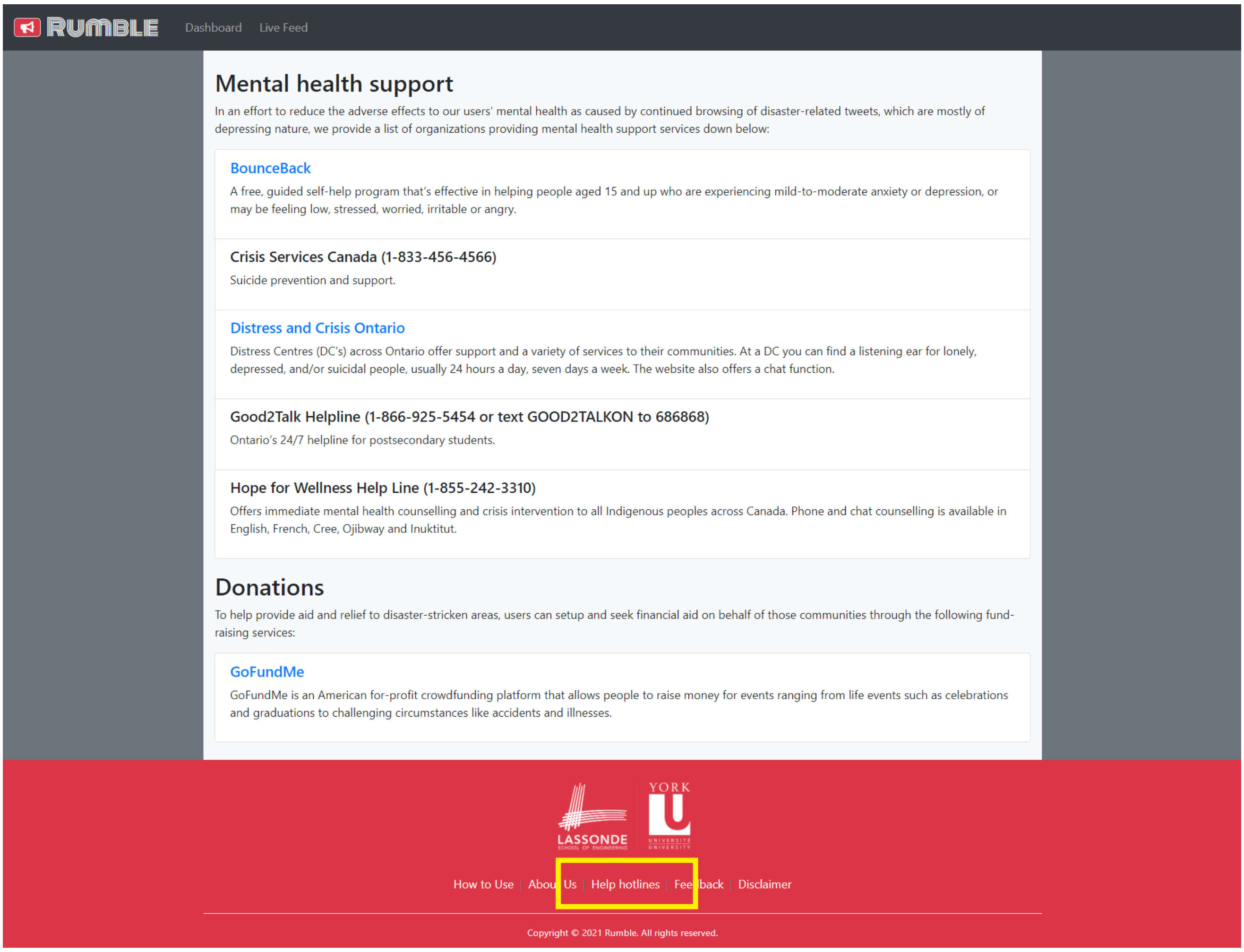
Steven O'Reilly @S_OReilly_USA

Figure 6: Rumble Live Feed Page

The Rumble Live Feed page includes a search bar for users to input disaster related phrases that they would like to retrieve Tweets for (e.g. in Figure 6 above, the user is retrieving Tweets containing the key phrase “volcanic eruption”). The feed then displays Tweets containing the search phrase, as well as the Tweet author’s user handle, phone, and verification status, and Tweet engagement data such as replies, retweets, and likes. A colour coded confidence bar and labelled percentage indicate the Tweet’s confidence level for being disaster “On-Topic”, and Tweets with a confidence level lower than 50% are colour blocked in red to indicate “Off-Topic”. The user can also sort Tweets ascending or descending based on either Timestamp or Confidence levels.

The Live Feed page also includes a “Filters” bar, giving users the ability to filter aggregated Tweet data by metrics such as Confidence, time frame, engagement metrics, and Tweet author’s verification status. These filters let the user parameterize the search based on their own needs and interests. For example, if a user wants to view only the aggregated Tweets that have a high confidence level of being On-Topic disaster related, they can set the Confidence filter to a high threshold such as 90%. By doing so, the Live Feed will display only those Tweets that have a minimum confidence level of 90%. The engagement filter selects Tweets based on the number of replies, likes and retweets that a disaster related Tweet has. The usefulness of this metric is to view the types of Tweets that garner a higher level of user engagement, and thus presumably have farther reach. The capacity to parameterize the Live Feed aggregate data allows users to easily tailor the aggregate data provided to more specifically suit their own personal use case.

Mental Health Support Page



Mental health support

In an effort to reduce the adverse effects to our users' mental health as caused by continued browsing of disaster-related tweets, which are mostly of depressing nature, we provide a list of organizations providing mental health support services down below:

BounceBack

A free, guided self-help program that's effective in helping people aged 15 and up who are experiencing mild-to-moderate anxiety or depression, or may be feeling low, stressed, worried, irritable or angry.

Crisis Services Canada (1-833-456-4566)

Suicide prevention and support.

Distress and Crisis Ontario

Distress Centres (DC's) across Ontario offer support and a variety of services to their communities. At a DC you can find a listening ear for lonely, depressed, and/or suicidal people, usually 24 hours a day, seven days a week. The website also offers a chat function.

Good2Talk Helpline (1-866-925-5454 or text GOOD2TALKON to 686868)

Ontario's 24/7 helpline for postsecondary students.

Hope for Wellness Help Line (1-855-242-3310)

Offers immediate mental health counselling and crisis intervention to all Indigenous peoples across Canada. Phone and chat counselling is available in English, French, Cree, Ojibway and Inuktitut.

Donations

To help provide aid and relief to disaster-stricken areas, users can setup and seek financial aid on behalf of those communities through the following fund-raising services:

GoFundMe

GoFundMe is an American for-profit crowdfunding platform that allows people to raise money for events ranging from life events such as celebrations and graduations to challenging circumstances like accidents and illnesses.

LASSONDE
SCHOOL OF ENGINEERING
YORK
U
UNIVERSITÉ
UNIVERSITY
How to Use | About Us | Help hotlines | Feedback | Disclaimer

Copyright © 2021 Rumble. All rights reserved.

Figure 7: Rumble Mental Health Support Page

The Mental Health Support Page provides the user with mental-health resources and GoFundMe links (currently in Ontario). In terms of sustainable development, Rumble seeks to enable its users with actionable resources as shown in Figure 7. The impact of this page is two-fold: first, because of the nature of disaster information, the help line page provides access to pre-emptive mental health support and resources for users themselves. Second, the aim of providing the links is such that users may in turn take action by the potential of further distributing (via Twitter engagement, or by other platforms) support links or donations to disaster events. The intention of Rumble is to promote the good health and well-being of society as a greater whole by the responsible dissemination of information, and so the promotion of important resources such as mental health support and charitable initiatives is essential to the message that Rumble seeks to convey.

Performance

The implementation of the as-built system uses a BERT classification model to label retrieved disaster keywords containing Tweets as being actually disaster related or not. The team found performance using the Kaggle dataset to train classification models well suited the scope of the project. Table 3 presents the results of the BERT and LSTM (best performing classification model as the Minimum Viable Product stage) models. The performance metrics were generated by testing the models using 10 stratified k-folds cross validation.

Model used	Accuracy of model predictions on the Kaggle test dataset	Precision	F1-score	Recall
LSTM + word embeddings (MVP)	0.813	0.822	0.767	0.720
BERT Classifier	0.829	0.869	0.781	0.710

Table 3: Performance of LSTM vs BERT Classifiers

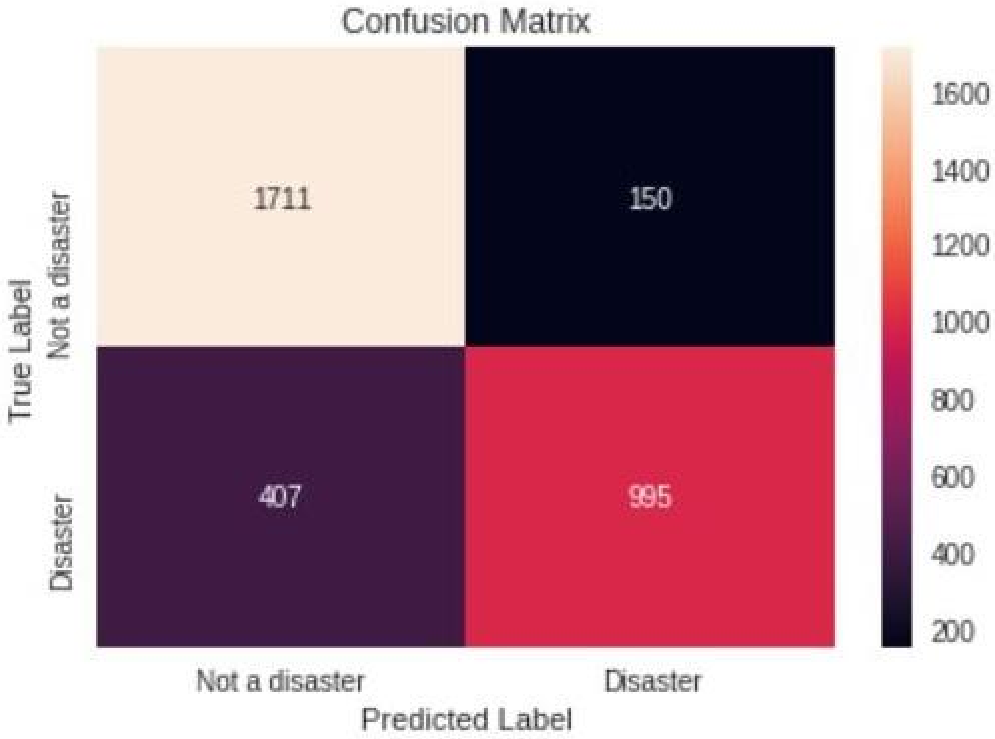


Figure 8: Confusion Matrix for BERT model

Comparing the best performing classifier from MVP release (LSTM + word embeddings) with the BERT model developed during Sprint 3, the BERT model produced better performance results. In particular, the significantly higher precision score of 0.869 indicates that the BERT model produces a lower number of false positives (Table 3). This is a key performance metric for the project, since a main objective is to present reliable true disaster Tweet labels on the website. The recall score indicates the percentage how many are being classified as true out of all the positives. In this case, the recall value is about the same as the MVP classifier. The F1-Score is slightly higher for the BERT model, but this is due to the precision score being much higher as the F1-score is a harmonic mean of the precision and recall scores. The accuracy was also higher, that was retrieved from the Kaggle test dataset.

The **confusion matrix**, seen in figure 8, shows how well the BERT Model performs on the test data where the true values are known. This shows that the model is able to distinguish between true positives (value: 995) and true negatives (value: 1711) very well. Especially as seen before with the precision score, it produces a low number of false positives (value: 150). This reinforces the fact that having a classifier that is good at labeling true disaster tweets is more beneficial for the user. So overall, the BERT model was an improvement on the previous model.

The limitations of the BERT model regarding **as-built design compliance analysis** are that it consumes more computational resources than other tested models. The decision to proceed with the BERT model is justified by its reputation as a state-of-the-art model in terms of Natural Language Processing, as well as its tested performance metrics with the provided Kaggle dataset. Because the purpose of the web application is to report whether a Tweet containing a disaster keyword is using the keyword in a context that is actually related to a disaster, the importance of metrics such as precision in producing a lower number of false positives is key to the performance of the model. As such, the tradeoff of higher runtime costs and computation time are offset by the benefits of BERT's better precision, F-score, recall and accuracy scores.

For failure tracking, the Github Issues mechanism is used to keep track of bug and error reporting. Detected bugs can be flagged and discussed between the team in the issue thread, and labels were tagged to the report to describe the issue and potential solutions. Team member(s) were assigned to the bug such that errors in the project can be systematically resolved. This mechanism allows for a detailed log of the bugs and errors that the team encountered throughout the incremental project Sprints. As a result, project version history was easily tracked and when similar bugs were encountered during Sprints, solutions and changes to the project were reviewed and resolved in an efficient manner.

Management Volume

Project Management Tools

The team's main communication methods were as follows:

Zoom: Video conferencing platform to hold our scrums and meetings with the supervisor, team meetings will be held on Zoom to have discussions about the project

Whatsapp: Group chat to provide weekly updates and discuss any minor/major issues faced by team members, as well as, communicating daily to bond better and developing a good team understanding

Google Docs/Slides and Microsoft Teams files: To share useful links and information, simultaneously work on shared documents, have targeted discussions as documentation for deliverables are produced by using comment/resolved threads, creating presentations for supervisor and peer reviews

Email: Professional platform to communicate with course directors and supervisors, asking questions, clarifying ideas, or getting help

Trello:

The team used Trello to manage tasks and keep track of product backlog, sprint backlog, tasks that are not started, in progress, and completed.

Sprints

A typical Sprint in this project would consist of:

- Sprint Planning (the team reflects on the Product Backlog, making necessary adjustments to it, and commits Product Backlog Items (PBIs) to the Sprint backlog)
    - Timeline (duration of a Sprint)
    - Roles- team members roles in a Sprint:
        - Product owner (Jessie Leung)
        - Scrum master (Binte Zehra)
        - Developer (Neena Govindhan, Paul Sison, Jonas Laya)
    - Sprint Goals (what needs to be achieved during a particular Sprint)
- Sprint Review
    - Completeness (successfulness of a Sprint), metrics used to measure completeness and efficiency:
        - Burndown chart (used by the team to track Sprint completion)
        - Velocity graph (used to outline the team's work output over the number of iterations or Sprints that have passed)
    - Challenges (difficulties faced during a sprint)
- Sprint Retrospective

- ○ Evaluation Metric (to measure the overall team satisfaction levels with a sprint, a Happiness graph was used to visually identify the Sprint events that team members thought worked effectively towards the product goals)
- ○ Process Related Issues (What went well in the Sprint and why? What can be improved for the following Sprints? What are actionable steps we can commit to improve for the next Sprint?)

Throughout the project, there were a total of 10 Sprints as summarized below:

Date	Sprint	Summary
02/11/2020-16/11/2020	**1.** Review performance of cross-validated recall, precision and f-scores on 'train' dataset.	● Implement an initial prediction model to generate predictions for the 'test' dataset using common classifiers. ● Implement a model with cross-validation technique to run on a 'train' dataset.
23/11/2020-07/12/2020	**2.** Implement a model using deep learning model (e.g. LSTM).	● Create tables comparing performance of different classification models. Use visualization tools (e.g. Confusion Matrix) to present classification results in a consumable form.
11/01/2021-24/01/2021	**3.** Reviewed MVP, tested and improved ML model	● Team review of MVP release and project goals, update PBIs for Alpha release ● Test and improve ML models including BERT (Bidirectional Encoder Representations)
25/01/2021-07/02/2021	**4.** Created website mockup and retrieved Stakeholder Feedback	● Create Disaster Tweet website mockup using Moqups ● Create a survey to test user interaction and get feedback on initial website mockup ● Reflect on the stakeholder feedback from the survey to prepare the website
08/02/2021-21/02/2021	**5.** Retrieved live Tweet stream and classified using ML models	● Obtain access keys to the Twitter API ● Curate a keyword list for the stream listener ● Configure Tweet database using PostgreSQL ● Process incoming Tweets and extract relevant information for analysis ● Store processed data to the database
22/03/2021-01/03/2021	**6.** Deployed the web application on a local server.	● Deploy a local web application using Dash ● Connect application to Tweet database and ML classification model ● Display live Tweets with prediction scores on the web application ● Display a map showing the tracked location of where the live Tweets were made, if geo-tagged

07/03/2021-14/03/2021	**7.** Deployed Website on cloud server and implement Live Feed with filters	● Deployed website and PostgreSQL Tweet database on Heroku ● Implemented live disaster related Twitter feed ● Implemented live feed filters (e.g. Confidence, Verified Users, Minimum Engagement)
15/03/2021-21/03/2021	**8.** Retrieved Stakeholder Feedback on Website	● Create survey to test the user interaction with the website ● Reflect on the results of the survey to plan for improvements
22/03/2021-31/03/2021	**9.** Improved Website and Geographic Search	● Improved the website based on the feedback from stakeholders acquired in sprint 4 ● Implemented the geographic search on the website
03/04/2021-11/04/2021	**10.** Implement Time Frame filter and Search by Keyword Function Tweet volume graphs and Improved website(based on final product survey) Add mental health and disaster relief resources, Analyzed SDGs, and Final Fixes	● Implemented another live feed filter (Time Frame) ● Implemented the search by keyword function ● Additional event page data visualization tools (implemented a tweet volume graph on the main page) ● Survey stakeholders (final product) for feedback and improved the final product ● Add mental health and disaster relief resources for users- GoFundMe links, hashtags for help line ● Analyzed SDGs and the social impact of this project ● Final fixes on the Website (i.e. formatting, labelling, etc.)

Table 4: Completed Sprints

Agile Process Review

The agile process was a great way to organize the tasks at hand. Organizing the tasks in the form of Sprints had helped with managing time and tracking the progress of the project. However, due to commitments in other courses, some Sprints took longer than expected. It was foreseen that it would be hard to maintain the weekly Sprints, as it could be seen in the Table 4, so from the beginning the Sprints were set to last around 2 weeks. Especially towards the end of the project with the final PBIs, enough time could not be allocated to completing them. In the future, realistic expectations of what can be achieved should be put in place, specifically when dealing with new topics, such as Machine Learning and web development in Dash, more time should have been allocated for the PBI.

Having designated roles, sometimes it was found to be easier to blend the roles and have discussions as a group when making decisions on what should go in the product backlog and sprint backlogs, instead of leaving it to just the product owner and scrum master to decide on.

For the sprint reviews and retrospectives, there was often not time to do all the graphs associated with them. So they were mostly done quickly and verbally to avoid spending too much time on that.

Instead of daily scrum meetings, it was found to be easier to have one 1-2 hour meeting a week and have constant discussions through WhatsApp, to keep track of progress, group members updating each other about tasks accomplished, and work together.

Final Budget

The project deployed a client-server style website such that users can view live natural disaster related news content. The user interface of the website directs and manages user requests loads from the server end of the website, which is responsible for scanning for new disaster-related Tweets, running them through the classification model, and returning the outputted news information to the user upon request. As such, user load management and database hosting are the key components of the project

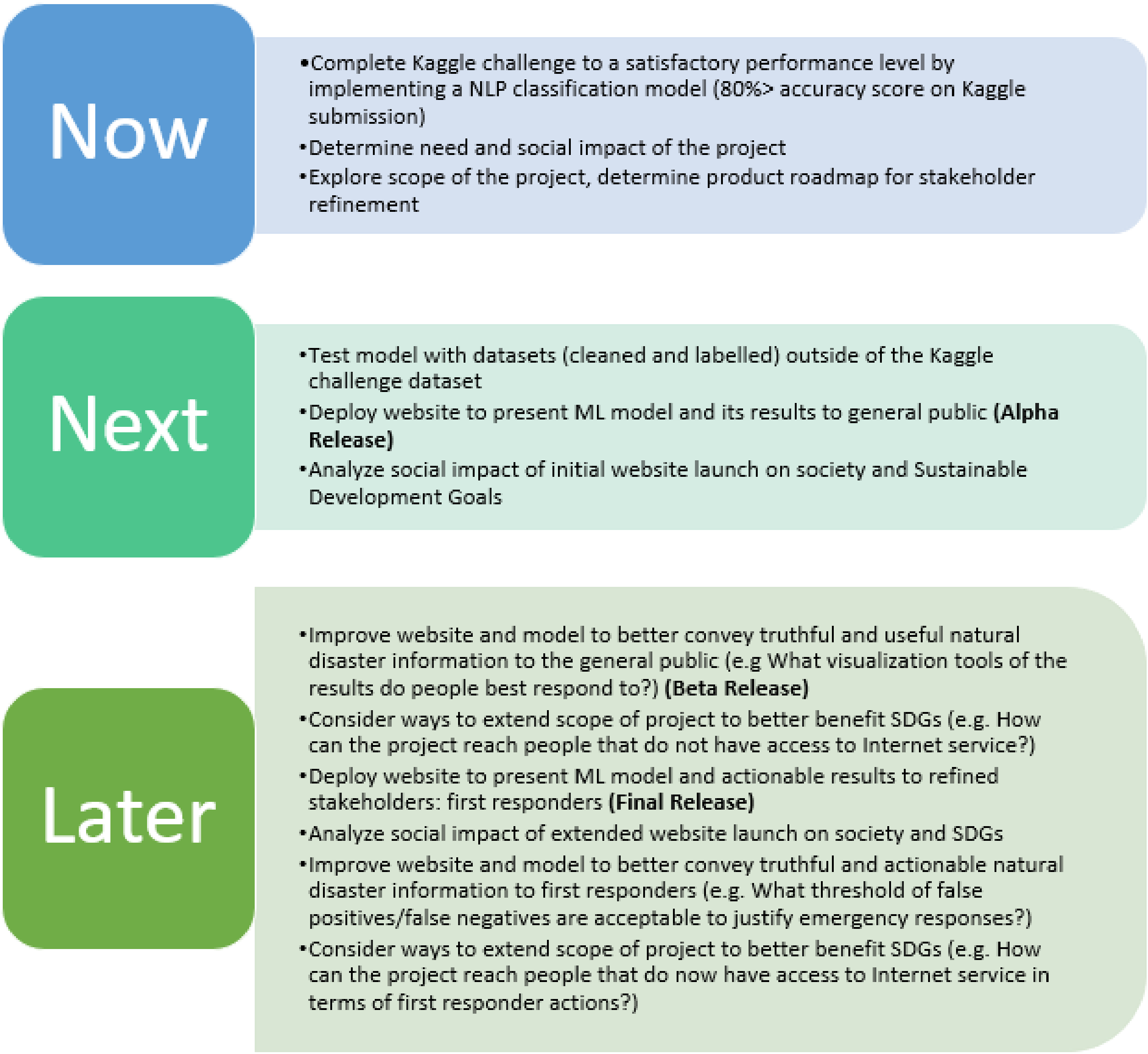
Given the $200 budget allocation for the ENG 4000 capstone, the team initially proposed that the budget will be allocated primarily for deploying the website using services such as Amazon Web Services (including services such as MongoDB, IC2 instance hosting, Docker Hub container management).

A detailed outline of the initially proposed budget allocation is as follows:

- $100 for the initial website deployment costs (e.g. AWS credits) targeted at the general public (requires larger amount of load management resources)
- $50 for project extension website deployment targeted at first responders (e.g. AWS)
- $50 reserved for miscellaneous or additional dataset acquisition costs (e.g. outsourcing of Tweet dataset gathering and labelling)

However, for the actual deployment, the free account on Heroku was found to suffice the needs of deploying the website and the database. So in turn we spent $0 as our final budget. Future costs would incur when deploying the Twitter Streamer to a cloud service as the file is too large for the free accounts on the cloud computing platforms available and to upgrade on the plans would incur a high cost for storing and running the streamer.

Project Schedule



Now

- •Complete Kaggle challenge to a satisfactory performance level by implementing a NLP classification model (80%> accuracy score on Kaggle submission)
- •Determine need and social impact of the project
- •Explore scope of the project, determine product roadmap for stakeholder refinement

Next

- •Test model with datasets (cleaned and labelled) outside of the Kaggle challenge dataset
- •Deploy website to present ML model and its results to general public **(Alpha Release)**
- •Analyze social impact of initial website launch on society and Sustainable Development Goals

Later

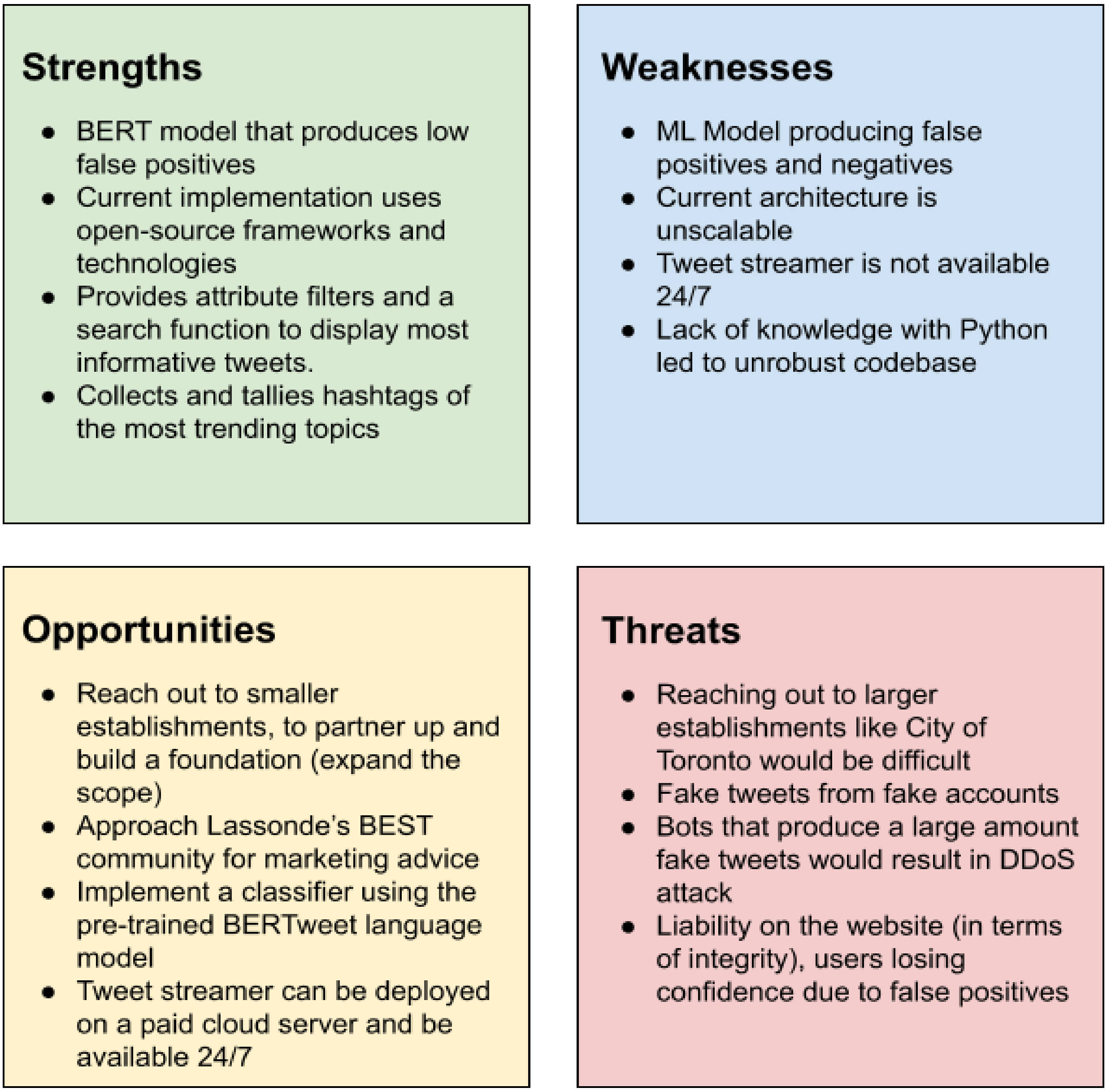
- •Improve website and model to better convey truthful and useful natural disaster information to the general public (e.g What visualization tools of the results do people best respond to?) **(Beta Release)**
- •Consider ways to extend scope of project to better benefit SDGs (e.g. How can the project reach people that do not have access to Internet service?)
- •Deploy website to present ML model and actionable results to refined stakeholders: first responders **(Final Release)**
- •Analyze social impact of extended website launch on society and SDGs
- •Improve website and model to better convey truthful and actionable natural disaster information to first responders (e.g. What threshold of false positives/false negatives are acceptable to justify emergency responses?)
- •Consider ways to extend scope of project to better benefit SDGs (e.g. How can the project reach people that do now have access to Internet service in terms of first responder actions?)

Figure 9: Now-Next-Later Framework from MVP

Provided above in Figure 9 was the initial plan of how the project will progress. But it is important to note that as the project progressed, there were some changes in terms of how it was planned versus how it flowed. Detailed overview of how the project components were delivered during Sprints is provided in Table 4. Additional details of what did not work well as the project progressed is given in Table 5. For the Final Release as mentioned above in Figure 9, in the 'Later' section, the team was supposed to deploy a website to present ML model and actionable results to refined stakeholders (First Responders); however, this did not work out as planned. The team decided to leave the scope to general users. It was important to make that decision to limit it just to the general public as they may benefit from the information provided on the website. If in the future, a high F1-Score is produced by the model, then this website would be able to successfully expand the scope of the project and refine stakeholders. A more thorough overview of the changes from the initial plan is discussed in the *Deviation from Plan* section under *Lesson Learned and Reflection* section.

Preliminary Business Case

The team employed **SWOT analysis** to evaluate the pros and cons of a potential decision. SWOT stands for strengths, weaknesses, opportunities, and threats and is a structured planning tool that aims to identify key internal and external factors influential to achieving a desired objective.



Strengths

- BERT model that produces low false positives
- Current implementation uses open-source frameworks and technologies
- Provides attribute filters and a search function to display most informative tweets.
- Collects and tallies hashtags of the most trending topics

Weaknesses

- ML Model producing false positives and negatives
- Current architecture is unscalable
- Tweet streamer is not available 24/7
- Lack of knowledge with Python led to unrobust codebase

Opportunities

- Reach out to smaller establishments, to partner up and build a foundation (expand the scope)
- Approach Lassonde's BEST community for marketing advice
- Implement a classifier using the pre-trained BERTweet language model
- Tweet streamer can be deployed on a paid cloud server and be available 24/7

Threats

- Reaching out to larger establishments like City of Toronto would be difficult
- Fake tweets from fake accounts
- Bots that produce a large amount fake tweets would result in DDoS attack
- Liability on the website (in terms of integrity), users losing confidence due to false positives

Figure 10: SWOT Analysis

Social Impact and SDG Goals

The dissemination of fake news on the Internet is currently hugely problematic in that it misleads the public in a far-reaching and lasting manner. This project seeks to automate truthful news extraction from social media posts in a timely manner, such that in times of emergency, actionable information can be accessed by the general public. The importance of this task is that in occurrences of natural disasters, immediate situational information can aid in the highly time-sensitive rescue operations of first responders. In disaster situations, even a marginal advantage could result in one or multiple lives being saved, or the prevention of catastrophic damages.

As an engineering capstone, one of the key project goals is to ensure the maintenance of the SDG goals. The following sections detail some of the ways that this project will benefit sustainable development.

Goal #3: Good Health and Well-Being. *Ensure healthy lives and promote well-being for all at all ages.*

By extracting reliable and useful natural disaster information from the mass amounts of information available on Twitter, the project aims to prevent the loss of life and increase the standard of living. The project mainly targets countries in which most of the population has Internet access readily available to them, but future extensions of the project could expand beyond the scope of the project defined in this document. The goals of this project and any project built off this one are such that healthy lives and the well-being of all people can be ensured through the dissemination of reliable, useful, and actionable crowd-sourced information.

Goal #8: *Decent Work and Economic Growth. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.*

As a potential future stakeholder, first responders would be a fundamental component in ensuring protection and service of society and the economy. As such, the project goal will be to provide first responders with timely and actionable emergency response information such that in times of natural disaster or crisis, disruption to human life and society can be minimized.

Goal #14 & #15: Life Below Water & Life on Land. *Conserve and sustainably use the oceans, seas and marine resources for sustainable development. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss.*

By extracting truthful natural disaster information in a timely manner, a key objective of the project is that in situations of natural disaster, destruction to natural resources and natural life both below water and on land can be prevented and minimized. The dissemination of truthful information can aid in efforts to protect and restore both human and natural ecosystems.

Goal #16: Peace, Justice, and Strong Institutions. *Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.*

The dissemination of truthful information from extractions of social media posts is an attempt to provide a service that is accountable and inclusive with the natural disaster information that it provides. A current weakness of the project scope is that its inclusivity is limited to where the general public have access to available Internet service. However, the project benefits attempt to overcome the malicious effects of fake news dissemination, and therefore aims to ensure accountable institutions and sustainable development goals.

Lesson Learned and Reflection

Deviations from the Plan

During the early stage of the project, the team's plan for the end product was to have a web service that utilizes the predictive model and live tweets to provide alerts to users regarding disasters happening. However, the team was caught by many surprises as the course progressed. Some Sprints took longer than expected due to time constraints and some components of the project demanded more time than the team anticipated. For this reason, the team deviated from the plan to provide alerts to the users because it would take a whole new set of learning aspects, such as topic modelling which was infeasible given the timeframe of the project. Instead, the decision was made to focus on improving how to relay disaster tweets, such as adding filters, graphs, geographic locations that would make it easier for users to get alerts from using our service, thus accomplishing the goal in a different manner.

The team originally sought to target first responders as the primary stakeholder as they would greatly benefit from the accurate information the end product can provide in response to disasters. However, due to time sensitivity and importance of their job, a very high F1-score was necessary to meet their needs. Providing them with a false positive and a false negative, such information can lead to wastage of precious resources and rescue operation's time. But further training of the model took a lot more time with marginal differences in results. Falling behind from the plan, the team decided to make the general public as their primary stakeholder so they can prioritize other aspects such as web development of the project. Thus, further training the model into a threshold that would refine the stakeholders into first responders became less priority.

Another deviation from the initial project plan included deployment and budgetary constraints. The original project timeline included deploying three components: the PostgreSQL database storing disaster keyword Tweets and Tweet metadata, the website server, and the Twitter streamer including BERT classification model. The Tweet database and website server were successfully deployed, thereby ensuring that the website is cloud-hosted and available to users on the Internet. However, it was found that hosting the BERT model on a cloud server would incur more budget costs than initially estimated, due to the size and runtime costs of the model.

First, the Tweet database is currently hosted on Heroku's free tier, which allocates a 10,000 row limit and 1GB storage capacity. Upgrades from the free tier are offered at pricing plans from $9/month to $16,000/month. Due to the nature of the project, which involves a disaster keyword Tweet streamer of many disaster related keywords, running the classification model in real-time would likely generate an immense amount of data to be stored. Furthermore, hosting the model itself on a cloud also incurred instance storage costs, especially due to the computational resource consumption of the BERT model.

The team found estimating the costs of deploying the Tweet streamer to be difficult, especially within the timeframe of the project, due to a lack of time available for scaling up and consumption testing. The costs of deploying the Tweet streamer has the potential to severely increase beyond the allocated project budget. As such, the decision was made to run the streamer locally to the Tweet database. This presents the opportunity to eliminate incurred costs while maintaining the real-time functionality and user Internet accessibility of the website. An extension of the project includes the possibility of selecting a deployment plan for the BERT model given an appropriate budget such that the classification model can be hosted on a cloud server.

Failure Report

Listed below are the main project failures that the team faced during the project and the lessons learned from those failures.

Main Project Failures	Lessons Learned from Failures
Failure to refine stakeholders to first responders	If the team decided to target first responders as a key stakeholder, it would be necessary to provide them with complete and accurate information. Thus, the F1-Score provided by the current model might not have been very useful for them. Sticking to the general public is safer rather than wasting precious resources and rescue operation time for first responders. This failure helped the team learn about the importance of designing a website that benefits limited stakeholders with accurate and useful information instead of wasting precious time of other stakeholders by expanding the project scope.
Failure to extract information from the Tweets to provide alerts using Topic Modelling	Implementing topic modelling to the project demanded more time than the team anticipated which led to deviation to the backup plan. The takeaway to this failure is the importance of time management and having a contingency plan. With an alternative plan ready, the team managed to implement a feature that would achieve the same goal as outlined above in the Deviations from the Plan section.
Failure to group related Tweets into events	Grouping related Tweets into events and grouping Tweets by keywords had very little difference in terms of accomplishment. Though, in terms of implementation, grouping by keywords can be done by a simple search tool which is a lot simpler than grouping related Tweets which suggests the need to use topic modelling. The difference in time and resource allocation can be spent on other aspects of the project. This failure conveyed that there are alternative paths that would save time and resources for more productivity.
Failure to deploy the Tweet streamer on the cloud	The team failed to deploy the Tweet streamer to the cloud due to its potential to incur higher cost than the allocated budget. As such, the decision was made to run the Tweet streamer locally to avoid such costs. Again, this showed the importance of having a risk mitigation plan as the team was able to limit the consequence.
Failure to develop an autonomous tool for maintenance	Developing an autonomous tool for maintenance was a low priority component planned for the latest stage of the project. The team prioritized the more important features that led to the failure which did not compromise the functionality of the product. This highlighted to the team the importance of assessing priorities during the early phase of the project.

Table 5: Project Failures

Lessons Learned

Below are the lessons that the team learned from each phase of the project that would be done differently, if given a choice to go back.

Project Phase	Lessons Learned
Phase 1: Agile Roadmap	Phase 1 of the project was mostly spent doing research and planning. What could have been done differently is balancing the learning aspect with the technical part of the project. Focusing on both the aspects and finding the right balance between the two could have made the process more efficient. This would have allowed the team to start early on the Kaggle challenge and this way, the team could experiment with additional datasets besides the Kaggle data.
Phase 2: Sprint Process Review	In Phase 2, along with the research and learning about the ML and NLP techniques, the team should have prioritized working with the Kaggle dataset and also implementing additional labelled datasets (e.g. from CrisisLex) to further enhance the model accuracy. This could have been split amongst the team.
Phase 3: MVP	During the Phase 3 of the project, the team was too focused on improving the performance of the model aiming for first responders as the primary stakeholder. But doing so yielded marginal differences in results. The team could have started working on the other aspects of the project such as web development at an earlier stage.
Phase 4: Alpha	There was a considerable amount of time between the MVP and the Alpha. But the amount of work done to the project was not proportional to the given time. The team took a lot of time in web development during this phase when it could have started earlier and could have added more features during the Alpha release.
Phase 5: Beta	In Phase 5 of the project, the team should have eliminated less crucial tasks. Spending time on trying new techniques did not work well for the team. As a result, for the future, taking out time to re-plan and reassess key components earlier on will leave time to enhance the end product for the purpose of visualization.
Phase 6: Final	For the Phase 6 of the project, the team should have been considerate of the time frame between Phase 5 and Phase 6. Eliminating low priority tasks during Phase 5 rather than pushing them to Phase 6, would have caused less stress and allowed time for final fixes on the website to make it look more visually appealing. Thus, in the future, realizing crucial tasks and just focusing on them would save time for any last minute fixes for the final product.

Table 6: Lessons Learned in Each Phase

Individual statements on ITPMetrics survey

Jessie Leung
The Peer Feedback given in the ITP Metric survey reinforced my sentiment that as a whole, my team was invested in working collaboratively to produce the best project that we could. My team constructively voiced both their appreciation and suggestions for how we could best work together. The feedback provided in the ITP Metric survey encouraged me to take note of my social and work responsibilities and how this might affect my contribution to the ENG 4000 capstone project. Resulting from this, I decided to make sure to set aside time to contribute for the project. My team members were also very supportive to this, and noted that the entire team would be accommodating to the varying schedules of each team member.

Binte Zehra
The ITP Metrics is a great tool to measure and reflect on my strengths and weaknesses as well as, how my team sees me versus how I judge myself. Specifically, the Peer Feedback helped me realize how my team appreciates my efforts and values my opinions. Their positive feedback regarding my performance, motivates me to work harder to achieve more and keep up the good work. In the future, if my team has any suggestions on how I could improve myself, I will take their suggestions and work towards achieving that goal. I want to work collaboratively with my team as well as enjoy this whole experience and come up with a project that can help the society in one way or the other.

Paul Sison
Prior to completing ITP Metric surveys, my suspicion was that peer feedback would only reinforce what I already knew were my weaknesses. But reading my teammates comments and their limitless support motivated me to work in a manner that would always see great benefit to the team. It was also surprising to learn that what I see as weakness can be appreciated by others.

Neena Govindhan
Doing the ITP Metrics Peer Feedback not only gave my peers a chance to give me feedback, but also allowed me to reflect on my own performance. The feedback I received from my peers has helped me to improve on the different aspects of the teamwork competencies, some of which I had also realized I needed to improve on. The positive feedback I had received made me realize what I was able to achieve and needed to maintain in order to be a good team member. However, I do believe there are always improvements to be made, i.e. check my blindspots, so I will strive to improve in all aspects and ask for my peers' feedback to keep improving.

Jonas Laya
The ITP Metric survey played a key role to my success in working with our team project. The feedback I received from the survey guided me on how to tackle the project with my team members. I greatly appreciated getting honest comments and suggestions from my team. With their constructive feedback, I was able to adjust myself accordingly to develop myself better with my team members. I am now able to communicate and contribute more with the team which created a more positive environment for us. The ITP Metric survey also created a great opportunity for me to show my appreciation and to voice out my suggestions to my team members.

Self-Evaluation

Criterion	Self Evaluation Ranking	Justification
Explain the importance of compliance with the Professional Engineers Acts and other relevant laws, regulations, intellectual property guidelines and contractual obligations and follow best practices	**Exceeding**	All Sections
Employ strategies for reflection, assessment and self-assessment of team goals and activities in multidisciplinary settings	**Exceeding**	Agile Process Review, Product Backlog, Sprints, Deviation from the Plan ,Individual statements on ITPMetrics survey
Adhere to written instructions in a professional context	**Exceeding**	All Sections
Evaluate critical information in reports and design documents	**Exceeding**	Design, Performance, Technical Volume
Appraise possible improvements in the problem solving process	**Exceeding**	Failure Report, Lesson Learned
Justify the strength and limitations of the solution and make recommendation for possible improvements	**Exceeding**	Preliminary Business Case, Failure Report, Lesson Learned

Table 7: Self-Evaluation

Appendix A: Glossary

- BERT: Bidirectional Encoder Representations from Transformers
- BoW: Bag of Words (feature extraction technique)
- CNN: Convolutional Neural Network
- DL: Deep Learning
- LDA: Latent Dirichlet Allocation (topic modelling)
- LSTM: Long Short Term Memory (Neural Network Classification Model)
- ML: Machine Learning
- MVP: Minimum Viable Product
- NLP: Natural Language Processing
- NLTK: Natural Language Toolkit
- PBI: Product Backlog Item
- RNN: Recurrent Neural Network
- SDG: Sustainable Development Goals
- SVM: Support Vector Machine (Non-Neural Classification Model)
- TF-IDF: Term Frequency-Inverse Document Frequency (feature extraction technique)
- TBC: To Be Completed

Appendix B: Disaster-Related Keywords

'accident', 'attack', 'blizzard', 'bombing', 'bushfire', 'casualties', 'catastrophe', 'collapse', 'collision', 'crash', 'criminal', 'cyclone', 'dead', 'death', 'debris', 'derail', 'destruction', 'devastation', 'disaster', 'drought', 'earthquake', 'emergency', 'epidemic', 'eruption', 'evacuate', 'evacuation', 'explosion', 'famine', 'fatal', 'fire', 'first responders', 'flood', 'forestfire', 'hail', 'hailstorm', 'hazard', 'heat wave', 'hijack', 'hostage', 'hurricane', 'injured', 'killed', 'killing', 'landslide', 'lava', 'lightning', 'mass shooting', 'massacre', 'meltdown', 'murder', 'naturaldisaster', 'oil spill', 'outbreak', 'pandemic', 'refugee', 'rescue', 'robbery', 'sandstorm', 'sinkhole', 'snowstorm', 'storm', 'structural failure', 'suicide bomb', 'survivor', 'terrorist', 'thunderstorm', 'tornado', 'tsunami', 'twister', 'typhoon', 'volcano', 'warzone', 'whirlwind', 'wildfire', 'windstorm', 'wounded', 'wreck'