ENG 4000: Alpha Gate 4

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

Date: March 5th, 2021

Team P

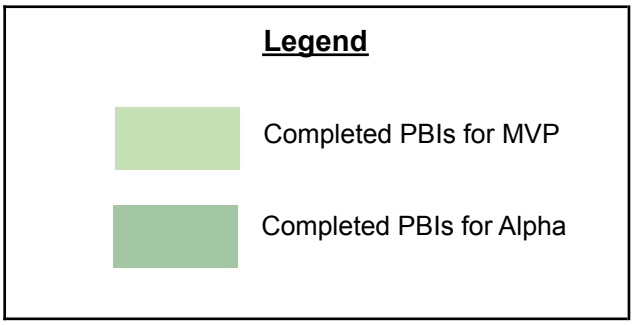
Binte Zehra	215624141
Neena Govindhan	212137600
Jessie Leung	215985948
Jonas Laya	214095715
Paul Sison	214447510

Table of Contents

List of Figures

List of Tables

Product Backlog



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~~	~~High~~
~~**16**~~	~~Clean and label new datasets~~	~~Medium~~
~~**17**~~	~~Test the new dataset on model~~	~~High~~
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	High
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)	High
28	Design methods to further improve the classification models based on new data	High
29	Develop tools to allow semi-autonomous maintenance of the website	High
30	Analyze SDG and the social impact of this project	High
31	Create a website Mock-up	High
32	Main page live feed table (Shows real and fake Tweets, make it visually appealing)	Medium
33	Group classified Tweets in to events based on keywords	High
34	Event pages (word clouds (relevant information), related Tweets, total number of Tweets, probability of event of each individual tweet)	High
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)	High
40	Survey stakeholders (final product) for feedback and improve final product	Medium
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 1: Product Backlog

Sprints

Sprint 3: Reviewed MVP, tested and improved ML model

Completed PBIs: 18
Tasks:

- Team review of MVP release and project goals, update PBIs for Alpha release
- Test and improve ML models including BERT (Bidirectional Encoder Representations)

During this sprint, the team evaluated the project goals as listed in the Minimum Viable Product Review document. Upon reviewing the performance metrics (including f1, precision and recall) of classification models built during the MVP stage, and discussing models to be tested such as the BERT model, the team found performance using the Kaggle dataset to train classification models to be well suited to the scope of the project. As a result, PBI #15-17 were removed from the project roadmap.

Upon building the BERT model and testing it using the same 10 stratified k-folds cross validation method used in the MVP review, the performance metrics of the ML models tested are summarized below:

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

Table 2: Performance of LSTM vs BERT Classifiers

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. This is a key performance metric for the project, since a main objective is to present reliable true disaster Tweet labels on the website. As such, the BERT model was chosen to be the classifier used in the development of the locally deployed website for Alpha release.

Sprint 4: Created website mockup and retrieved Stakeholder Feedback

Completed PBIs: 24, 31
Tasks:

- 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

During this sprint, the team created website mockups to visualize features required for the actual website. The mockup was based on the system Use Case diagram as shown in Figure 13. It included three page components and are shown in the figures below:

- Main landing page: includes a live stream of disaster Tweets, extracted disaster information grouped into “Events”, Tweet geolocation information displayed on a map
- Search results page: includes Tweets related to the keyword or location searched.
- Event details page: includes miscellaneous information about the event such as related Tweets, location, predicted probability and relevant information.

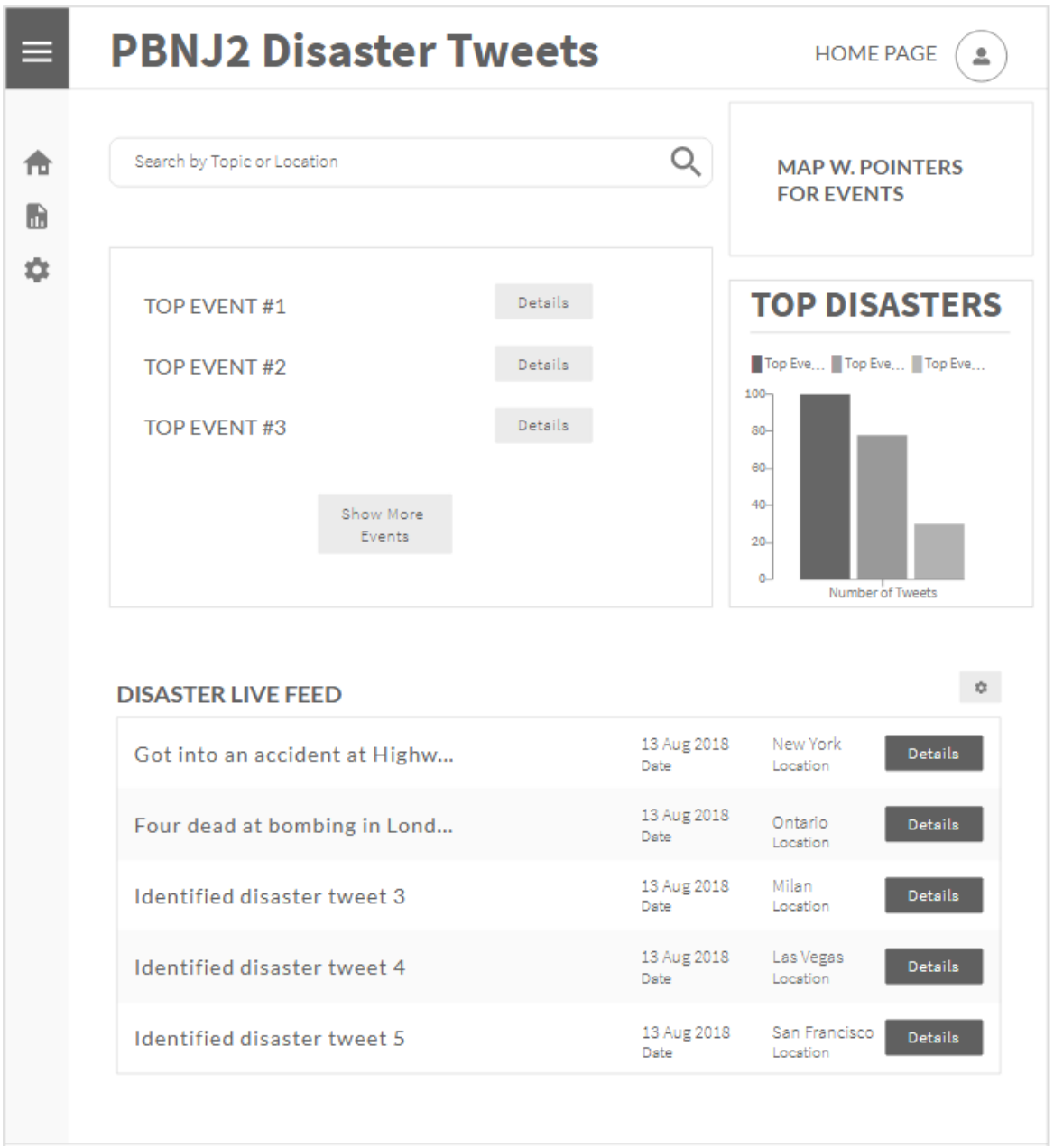


Figure 1: Main Page Mockup

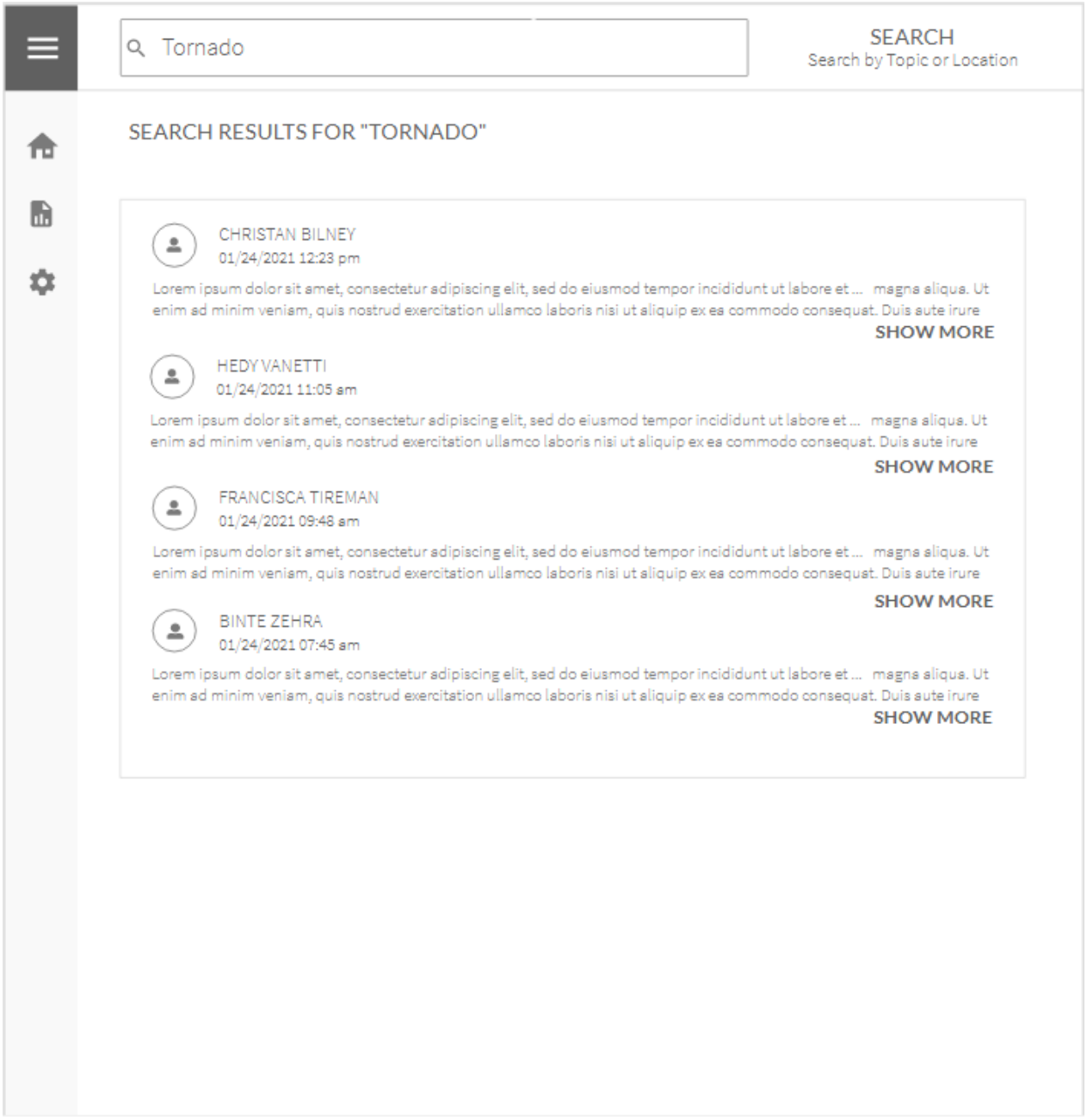


Figure 2: Search Results Page Mockup

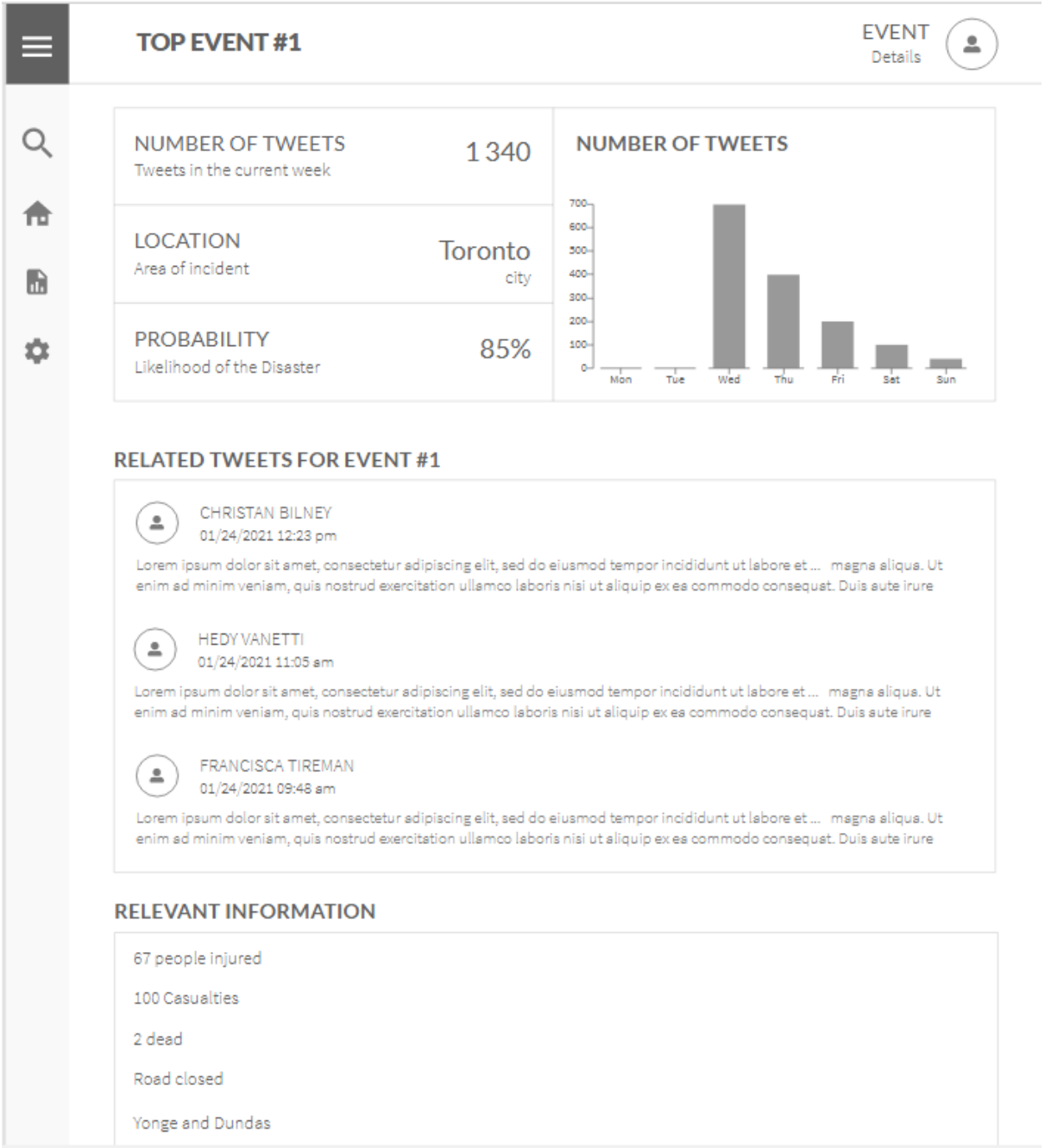


Figure 3: Event Details Page Mockup

To elicit stakeholder feedback on the project at its current stage, the mockups were presented to a sample group of the target stakeholders (general public) and a survey was distributed such that user response to the mockup website user interface could be observed. Some evaluation parameters included in the survey are summarized by the following:

- Is the user interface visually appealing?
- Is the user interface user friendly?
- What can be improved?
- How likely are you going to use our website?

For Home page:

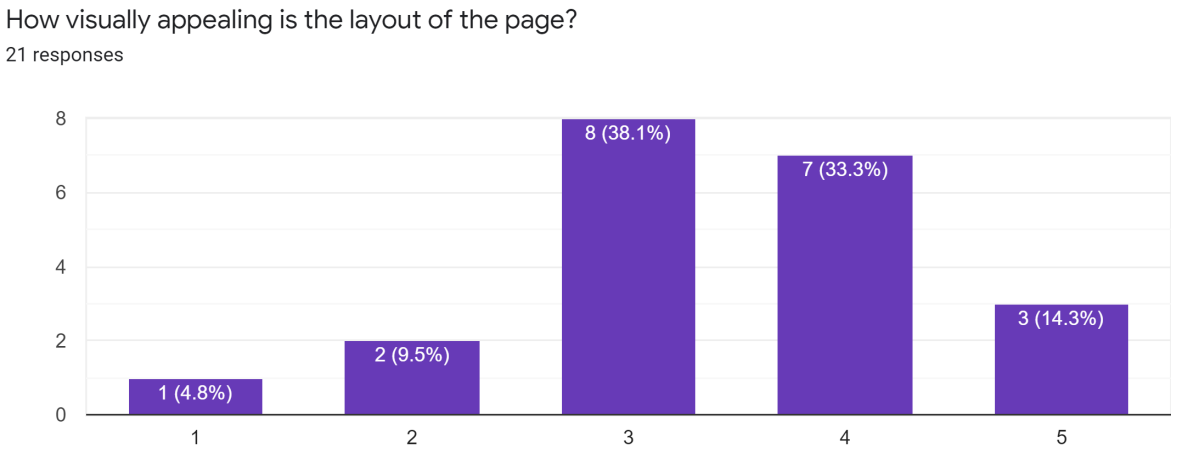


Figure 4: Home Page Visual Survey Results

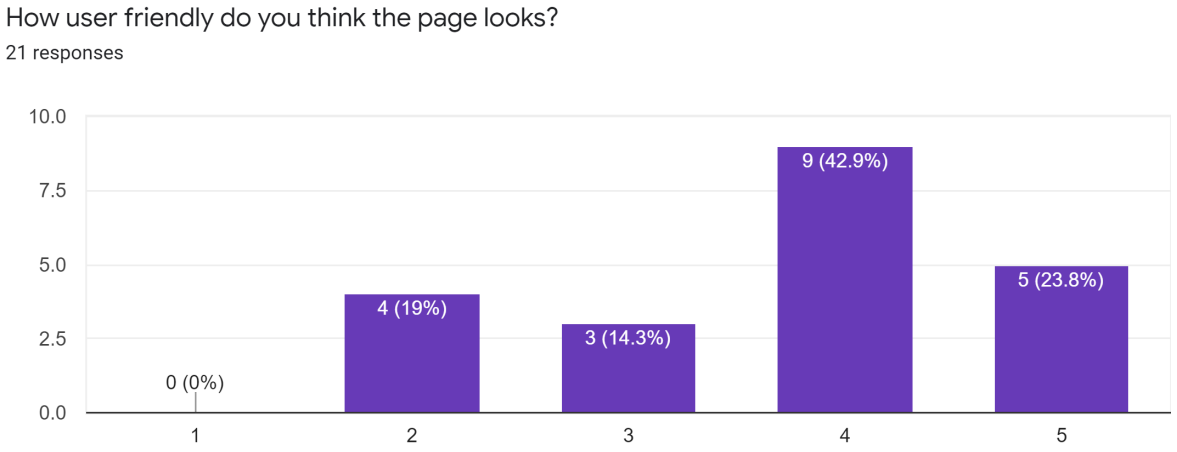


Figure 5: Home Page User-Friendliness Survey Results

For Events Page:

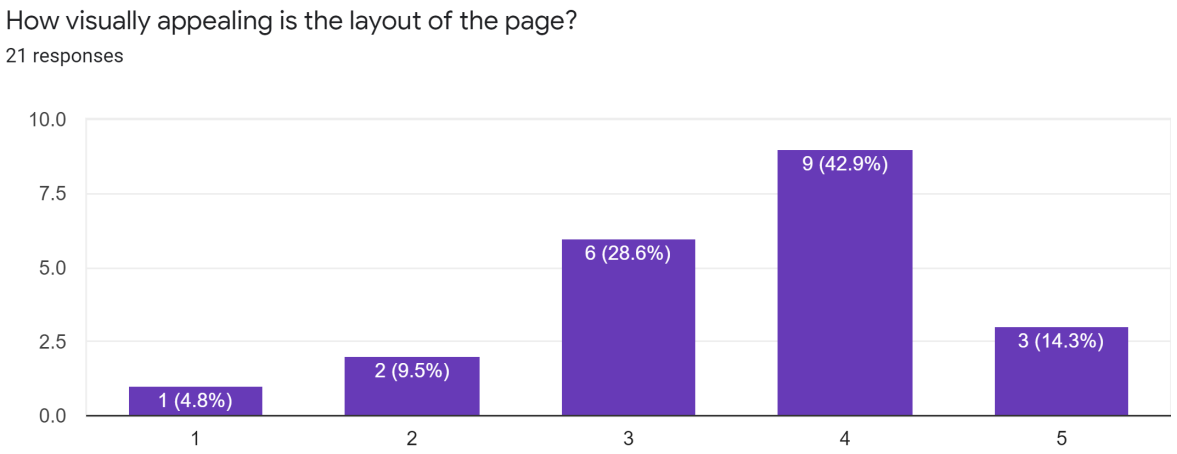


Figure 6: Events Page Visual Survey Results

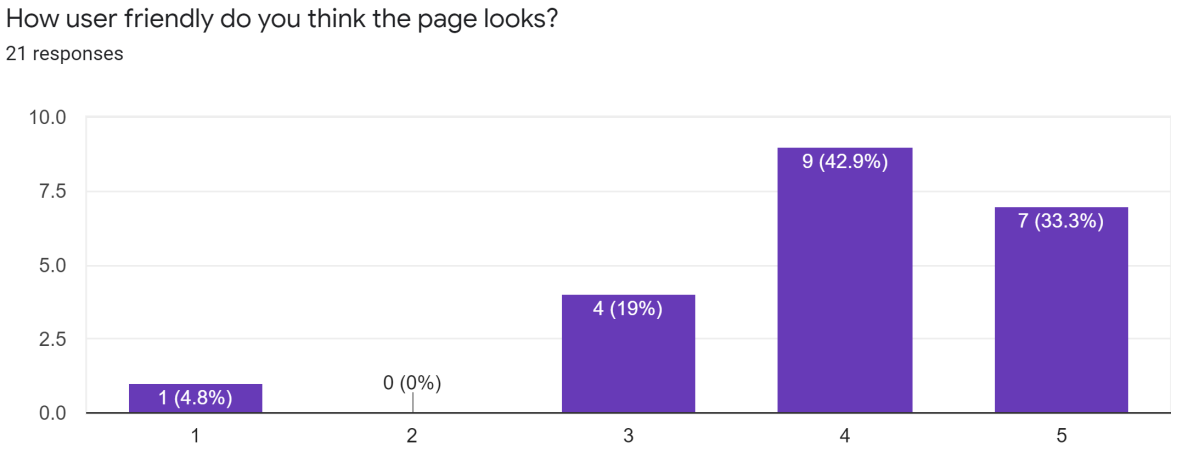


Figure 7: Events Page User-Friendliness Survey Results

For Search Results Page:

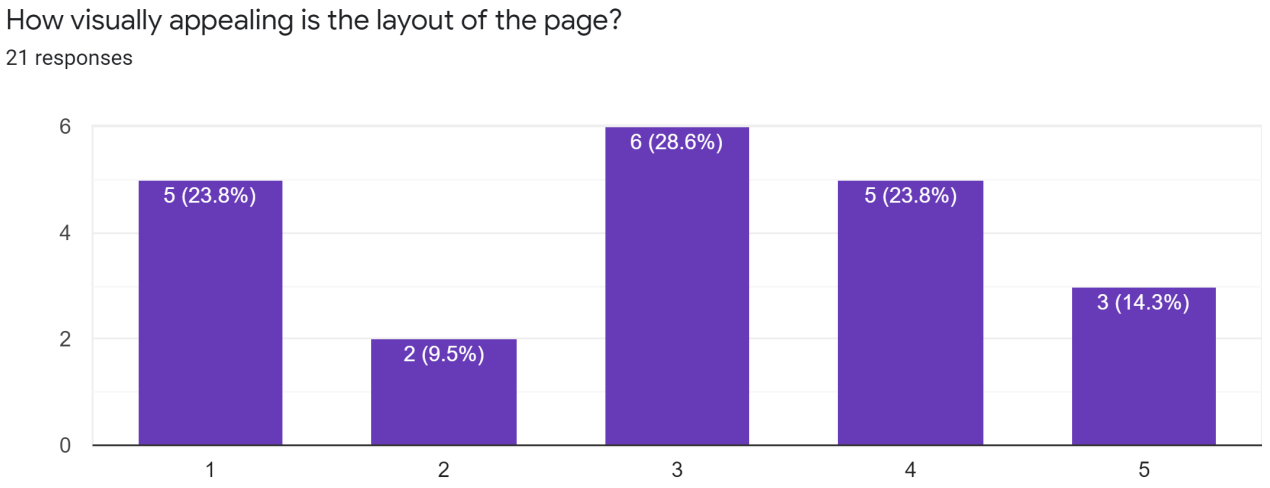


Figure 8: Search Results Page Visual Survey Results

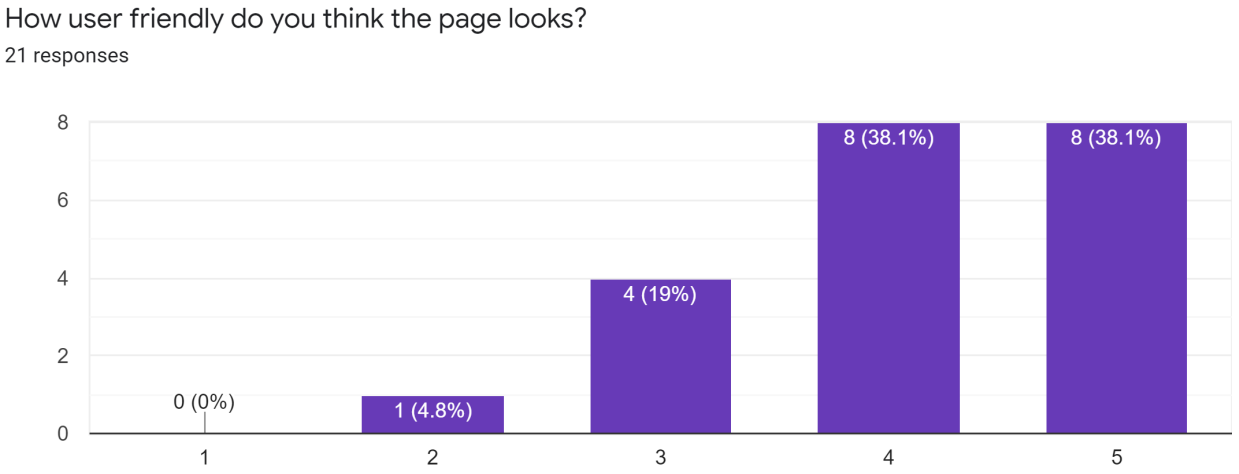


Figure 9: Search Results Page User-Friendliness Survey Results

Overall Website:

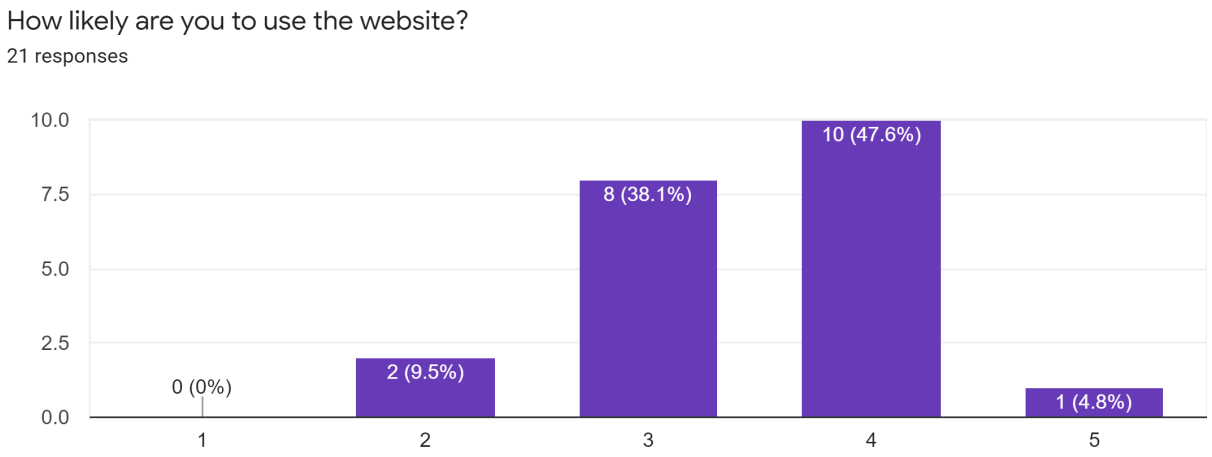


Figure 10: Website Usage Survey Results

Stakeholder Feedback on the how the website can be improved:

Search page looks empty
There is a lot going on the main page, needs more colors
Adding colour
The main page looks a little cluttered. Color might be a good thing to add. What is the map w. pointers?
Nothing! What's the point of this website??
Make it more interesting!
I DON'T GET IT!!
It looks great!
Designs overall look a little boring and how is this website useful for me?
Search results page does not look good
Needs work on the overall looks
I like the concept and the simplicity. Could be more colourful though.
Layout could be better.
The layout looks okay. Could have better graphics. What are the buttons for on the side of the page?
Looks good but needs some rework on the design
Color is so boring so you might want to work on that
It looks okay but I think because these pictures are in black and white it makes it less appealing
For an improvement I would say I don't think it's visually appealing. For the text it's concise but on the events page there's a lot of white space that could be optimized better.
Pages are black and white. Definitely needs improvement
Better graphics and should show verified tweets.
Provide more real estate for the map if that is a significant functionality to the app.

Upon reviewing the survey results, some general responses found to be common between multiple potential users are summarized in the table below. During this sprint, the team discussed the stakeholder feedback from the mockup surveys and elicited potential opportunities for improvement from user suggestions.

User Response	**Future Improvements**
Website interface is user friendly and easy to navigate for most users	Keeping a simple layout for the website should keep it user friendly and easy to navigate.
Some users were unsure of the purpose of the website	Website can include a small explanation or help page for what a “disaster Tweet” is and a simple summary of how the classification and prediction data that is presented in the website is produced.
Use of more colours will make the website more visually appealing	Colour coding can make the website more interesting and informative (e.g. coding “true” disasters Tweets in green and “fake” disaster Tweets in red)
Unsure of what the navigation symbols on the side of the website are indicating	Navigation bar can be more clearly embedded in the top of the website layout and consist of clear text to indicate their purpose
Use of data visualization tools such as graphs and/or maps make the website look more professional and scientific	Data visualization tools created using frameworks such as Dash can be used to present data to users in a way that is informative and interesting to users

Table 3: Potential Future Improvements from User Feedback

Sprint 5: Retrieve live Tweet stream and classify using ML models

Completed PBIs: 22, 23

Tasks:

- 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

During this sprint, the team obtained an access key to the Twitter API by submitting a request through Twitter developers account. Using Tweepy and the Twitter API, a Tweet streamer is built to retrieve Tweets that contain the following disaster related keywords: ‘natural disaster’, ‘storm’, ‘typhoon’, ‘earthquake’, ‘tsunami’, and ‘volcano’. Information

extracted by the streamer includes data such as the full text of the Tweet, user's Twitter handle, geotagged location of the Tweet (if it exists), Likes, Retweets, and other information.

The streamed Tweets are then classified by the BERT model developed in Sprint 3 and a confidence metric is produced (indicating the model's confidence that the Tweet is truly disaster related). Along with the Twitter data, this classification label is then stored into a PostgreSQL database.

Sprint 6: Deployed a local web application

Completed PBIs: 19, 21
Tasks:

- 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

➔ System Configuration and Failure Tracking

As an initial step to web deployment, the web application is first developed and deployed on the team's local systems. The web application is built using a Dash framework, such that in future releases data visualization tools such as tables or graphs can be easily implemented into the design. The team decided on a Client-Server architecture for the high-level design of the application. The main components of the system are described by the figure below.

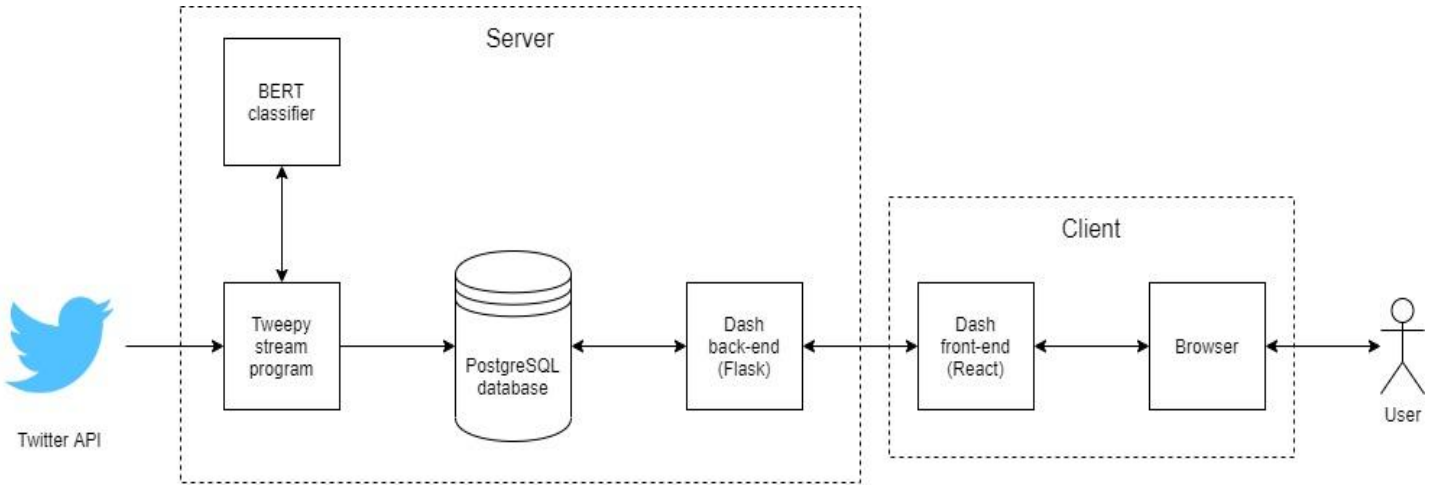


Figure 11: System Diagram for the Web Application

Following the above architecture design, components of the locally deployed web application were built. The web application was connected to the PostgreSQL database implemented in Sprint 5, which streams disaster related Tweets and labels them using the BERT classification model before storing them. The web application retrieves this data, and then presents the data on the main landing page in two main components:

- **Live Twitter Feed:** Live stream of Tweets including disaster keywords (e.g. earthquake, storm, etc.) that are classified as disaster related. User handle, full text of the Tweet, and engagement information (Replies, Likes and Retweets) are displayed. The confidence level of the Tweet being disaster related produced by the Bert model is shown, and the Tweet container is colour coded based on the confidence level (e.g. 65% confidence that Tweet is disaster related is colour coded Yellow, 85% confidence that Tweet is disaster related is colour coded Green)

- **Tweet Indicator Map:** A geographic map with flagged points indicating a geotagged disaster related Tweet. If the user hovers their mouse over the flag, geolocation coordinates and text of the Tweet are displayed in a pop-up container.

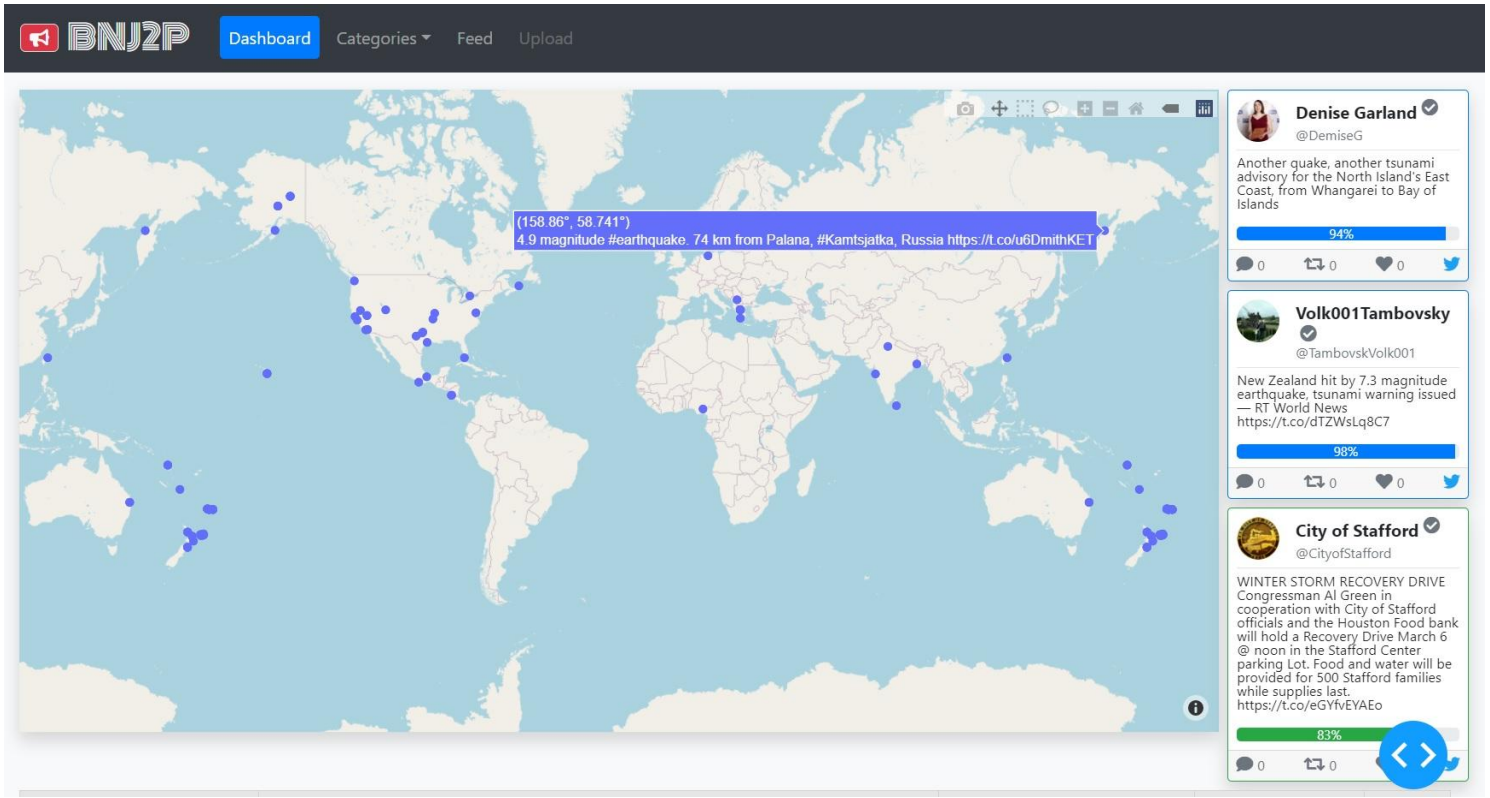


Figure 12: Locally Deployed Web Application

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 can be tagged to the report to describe the issue and potential solutions. Team member(s) can be 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 can be easily tracked and if similar bugs are encountered in future sprints, solutions and changes to the project can be reviewed and resolved in an efficient manner.

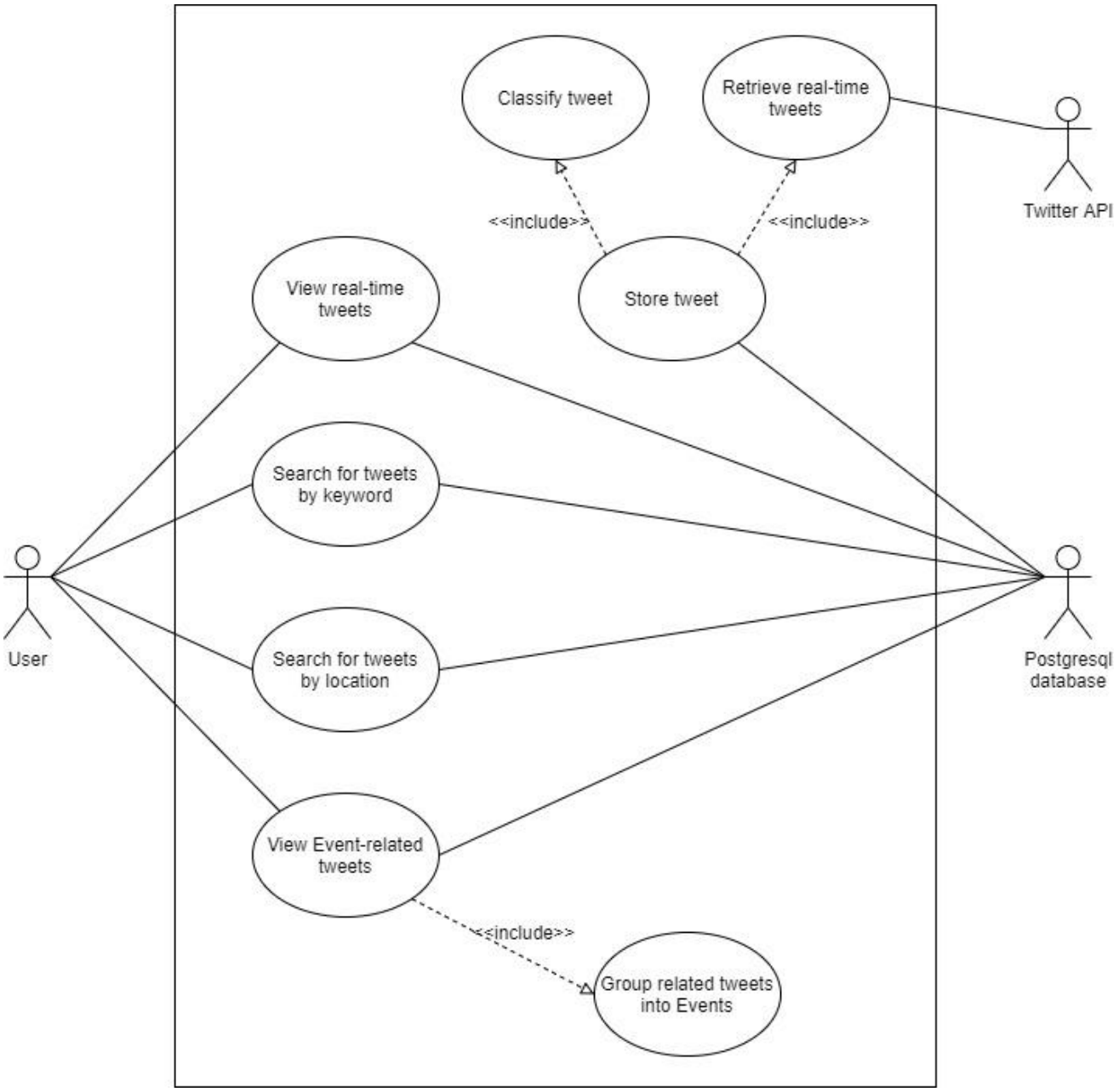


Figure 13: System Use Case Diagram

The use case diagram above 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.

System Issues

MVP (Issues)	Alpha (Solution)
Simple ML classification models reached the target of 80% accuracy score on the Kaggle challenge. Alpha release goal is to develop a model with improved performance.	The BERT model built and trained on the Kaggle dataset produced better performance metrics as presented in Table 2.
Presenting the data of the ML classification models in a visually appealing manner to the stakeholders	Developed a locally deployed website to showcase data. The website was deployed using Dash. It presents a real-time feed of Tweets containing disaster keywords and a confidence label generated by the BERT model that describes the likelihood of the Tweet being truly disaster-related.
Unable to retrieve real time Tweets	Live Twitter feed table made using Twitter Developer API and Tweepy streamer
Tracking location of the Tweets and displaying it to the users	Consists of a map, pinpointing location of real-time disaster related Tweets
Classification model was not used outside the Kaggle Dataset	Model was combined with the Tweepy streamer to classify Twitter live feeds

Table 4: MVP Issues and Provided Solution in Alpha

Path to Beta and Final Product Release:

Beta Release	Final Product Release
● Event grouping (keywords)***** ● Event pages (Top 3 events)*****: ○ Word clouds (relevant information) ○ Related Tweets ○ Total number of Tweets ○ Probability of event of each individual tweet ● Geographic search ● Survey (functionality and visual appearance) ● Deployed website ● Search Page (Static result table with a refresh button) ○ keyword search ○ Display related Tweets based on engagement (likes, quotes, retweets) and confidence of >90% ● Main page live feed table (Shows real and fake Tweets, make it visually appealing)	● Final fixes on the website ● Additional event page data visualization tools (Tweet volume graphs)***** ● Survey stakeholders (final product) for feedback and improve final product ● Disaster relief and mental health resources.******

Table 5: Path to Beta and Final Release

***Event Page**

The website will consist of event pages. The main page will display top three events and each event is grouped together using the same keywords of a disaster related tweet. When a user clicks on a top event on the main page, they are directed to the event page. The event page will have related Tweets as well as the probability of an event. It will show the total number of Tweets and a word cloud, showing all the relevant/useful information. As for the final release, the event page will have a tweet volume graph as a visualization tool to further enhance it.

****Disaster relief and mental health resources**

This section 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.

ITP Metrics Reflection:

Individual statements on ITPMetrics survey (Peer feedback)

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. The suggestions that my team made regarding how we could improve as a team and work together better to design a project that leaves an impact on the society. 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.

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, so I will strive to improve in all aspects and ask for my peers' feedback to keep improving.

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.

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 highlighted my strengths and weaknesses which decided my role in the team. As for the peer feedback, 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 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.

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.

Team statement on ITPMetrics survey

As per our ITP Metrics, Team Dynamic report, as a whole we work well as a team. In terms of Communicate, compared to the Gate 1 report, now our Communicate score had increased. Because we are following the Agile project management approach we have been in more contact in terms of scrum meetings and we have more defined roles that are still flexible. We also have gotten to know one another much better, that we are more comfortable to be more open with one another.

In terms of Adapt, again our scores have increased in this gate. This is because prior to Gate 1, we were splitting up the tasks and putting it together and going over what needs to be changed at the end. But now there is more communication because of the scrum meeting and more in tune with what each other is doing. Still there are some days where time management is still an issue, because of other courses and personal lives, but with the planning we are more on track than before. Even if certain tasks don't go exactly as planned, we planned some buffer time so that there is time to catch up with the tasks.

In the Relate section, as a team we have been good at contributing to work equally and have had a positive environment where no conflicts have arisen as such. That hasn't changed much since. In terms of conflict, we hardly have major ones, we usually talk it out and get everyone's opinion and if someone is wrong, they would openly admit it. As a team we try to maintain a positive energy and environment for work. And we have built that trust in one another to do that work.

In the Educate section, since this was a new topic for all of us, this has been by far a very important part of the process. Since we all have different learning styles, we leveraged on team members that were able to grasp material much faster to get certain tasks done which we planned during the scrum meeting, but we all continue to learn and contribute to the progress of our project. Especially, when we started the sprint, we focused on learning specific topics such as learning the different classification models we want to use for the project. Since this was more specific this task was easier to accomplish, than before where we did more general research about ML and NLP. But we do understand that performing the general research has enabled us to understand the specific topics. We still continue to be open to learning and especially now will be focusing on the social aspect.