**Team Name:** Danger Alert

**Team Members:**

- Shadi Nourriz
- Drew Remmenga
- Devan Neil

**Section Number and Group Number:**

- Section , Group 15

**Name of the Kaggle Project Chosen:**

- Natural Language Processing with Disaster Tweets

---

## Proposal Page

### Problem Statement

**What is the problem you hope to solve via your ML product?** We aim to predict whether a tweet is about a real disaster or not. This can help emergency response teams quickly identify and prioritize tweets that require immediate action.

**Who will want to use the product and why?** Emergency response teams, disaster relief organizations, and government agencies will benefit from this product to quickly filter and respond to critical information during emergencies. Social media monitoring platforms can also use this to enhance their disaster response capabilities.

**What are current solutions to the problem (if any), and why might your ML product be an improvement?** Current solutions often involve manual monitoring or simple keyword filtering, which can be inefficient and inaccurate. Our ML product will leverage advanced Natural Language Processing (NLP) techniques to provide more accurate and automated detection of disaster-related tweets, significantly improving response times and resource allocation.

### Problem Solution

**What class(es) of model do you expect to try/use?** We plan to explore various NLP models, including Logistic Regression, Support Vector Machines (SVM), Random Forests, and advanced deep learning models such as Long Short-Term Memory (LSTM) networks and BERT (Bidirectional Encoder Representations from Transformers).

**How will you quantitatively score the performance of your model?** Performance will be measured using metrics such as accuracy, precision, recall, and F1-score. These metrics will

help evaluate the model's ability to correctly identify disaster-related tweets while minimizing false positives and false negatives.

**How well do you believe the model needs to perform for it to be useful?** For the model to be valuable, we aim for an accuracy above 85%, with high precision and recall scores (above 0.80) to ensure reliable identification of disaster-related tweets.

**What performance do you guess the model might achieve?** Based on initial explorations, we anticipate achieving an accuracy around 87%, with precision and recall scores each around 0.85, providing reliable detection of disaster-related tweets.

**The Data**

**A short description of your data (data source, size, structure, features, etc.):** The dataset is sourced from Kaggle and contains 10,000 tweets labeled as disaster or not. Each tweet is represented by text data, and additional features such as keyword, location, and the presence of hashtags may be included.

**How will you use them for training and testing?** The data will be split into training and testing sets, with 80% used for training and 20% for testing. Preprocessing steps will include tokenization, stop words removal, stemming, and converting text data into numerical representations using techniques like TF-IDF or word embeddings. Cross-validation will be used to ensure model robustness.

---

# Additional Info

## Figures and Plots

**Data Distribution Plot:** A plot showing the distribution of disaster-related and non-disaster-related tweets in the dataset.

**Word Cloud:** Word clouds for both disaster-related and non-disaster-related tweets to visualize the most common words in each category.

**Model Performance Plot:** Graphs comparing accuracy, precision, recall, and F1-score across different models (Logistic Regression, SVM, Random Forest, LSTM, BERT) to visualize model performance.

## Timeline

| Milestone | Target Date |
|---|---|
| Construction of initial model (untrained) | July 15, 2024 |
| Training of initial model and performance assessment | July 22, 2024 |

| | |
|---|---|
| Training of revised model(s) and performance assessment | July 24, 2024 |
| Project submission on Kaggle | July 25, 2024 |

**References:**

1. Kaggle Dataset: Natural Language Processing with Disaster Tweets
2. Scikit-learn Documentation
3. Research papers on NLP techniques and models