CSC4260/Advanced Data Science Project:
Natural Language Processing with Disaster Tweets
Project Plan and Tentative Timeline

Project Members: Sharon Colson, Caleb Smith, Thomas Robertson, Tania Flores

1. Project Overview
- **Objective:** Develop a Natural Language Processing model to predict which Tweets are about real disasters and which ones are not by evaluating accuracy using F1 Score.
- **Key Deliverables:** Project Report, Poster, NLP Model with an F1 score of 0.85 or higher

2. Agile Approach
- **Framework:** Scrum (2-week sprints)
- **Process:**
  - Sprint Planning → Development → Peer Programming → Sprint Review → Retrospective
  - Deliverables prioritized in a backlog and refined each sprint.
  - Continuous integration and iterative model improvements.

3. Tentative Timeline

Iteration	Duration	Key Activities	Deliverables
Iteration 1	Week 1-2	Define scope, gather data sources, initial data exploration	Data sources identified, EDA report
Iteration 2	Week 3-4	Data cleaning, feature engineering, baseline model	Clean dataset, baseline model
Iteration 3	Week 5-6	Model experimentation, hyperparameter tuning	Improved ML model
Iteration 4	Week 7-8	Model evaluation, deployment strategy	Production-ready model
Iteration 5	Week 9-10	Testing, final refinements, stakeholder demo	Final model
Iteration 6	Week 11-12	Final model documentation and poster	

4. Continuous Monitoring & Feedback
- Regular team check-ins for adjustments once a week; every Tuesday night on virtual team meetings as well as provided class time from our professor.
- Performance tracking post-deployment.

5. Models
We will be looking at standard Binary Classification models such as Logistic Regression, Binary Trees, KNN, Support Vector Machines. Regarding Natural Language Processing, we will also utilize various supportive transformer models utilizing DistilBert, RoBERTA, BERT, and other transformative models in terms of that. Additionally we will look at adding some type of outlier detection.

6. Assignments
With twice weekly meetings, the team intends to work through each phase of the project together. However, the lead on each task will be divided as follows:
- Project Planning and Initial EDA - Tania
- Project Reporting - Sharon
- Base Models - Caleb
- Final Models/Hypertuning - Thomas