Capstone III Project Proposal

Problem Statement:

1. How can we build an NLP model that can classify any given tweet as about a disaster or not while maintaining 90% accuracy or better?

Context:

With the rise of social media comes tons of data on how people interact with one another through use of the internet. While it is easy for a human to understand the intent of a post on social media, it is much harder for a machine. This is mainly due to use of sarcasm and buzzwords being used under many different contexts. It is for these reasons that I am interested in looking into tweets containing disaster keywords and building an NLP model that can effectively classify untrained tweets as either about a real disaster or not. This is extremely valuable to organizations that provide disaster relief as they can more efficiently identify real disasters from fake ones. This same project structure could be used in conjunction with other data to detect fraud, perform sentiment analysis, and has many more applications.

Criteria for Success:

1. Correctly classifying tweets as about a disaster or not (aim for 90%+ model accuracy).

Scope of Solution Space:

The focus will be put on building the most accurate models possible so that tweet classifications are reliable and useful. The data we will use contains Twitter data from January 14th, 2020 and should ideally be retrained periodically to include new disasters.

Constraints:

If our best model cannot support 90% accuracy feasibly we will have to adjust our goals to align with what our data realistically tells us.

Stakeholders:

Primary: Disaster relief organizations

Data Sources:

Data source link: https://www.kaggle.com/datasets/vstepanenko/disaster-tweets

All of the data was scraped from Twitter. The tweets were collected on January 14th, 2020. Topics people were tweeting about include the eruption of Taal Volcano in Batangas, Philippines, COVID-19, bushfires in Australia, and teh Iran downing of airplane flight PS752. The data consists of 1 table, *tweets.csv.*

Deliverables:

1. Notebooks for all of the steps throughout the project.
2. A project report going over the process and key findings.
3. A slide deck going over the process and key findings.