Real or Fake? NLP with Disaster Tweets - Project Idea

Susanth Dasari
Computer Science
Georgia State University
Atlanta, Georgia
sdasari3@student.gsu.edu

Sameera Turupu
Computer Science
Georgia State University
Atlanta, Georgia
sturupu1@student.gsu.edu

1   Problem

1.1   Introduction

Twitter is one of the most popular social media platforms and is used by millions of users around the globe to project their view into the world about a wide range of topics which include generic topics such as movies and music, and also sensitive topics such as politics and disasters. Twitter has become an important communication channel in times of emergency given the ease of usage of smart phones. In such times, agencies are interested in monitoring tweets to identify such disasters.

When we are talking about identifying disasters through a tweet, there are a lot of obstacles to consider. In this project *we aim to classify **disastrous tweets**, as opposed to tweets on irrelevant topics*. A particular challenge with this task is that both classes contain the same choice of words, which can be differentiated by using the context of the sentence.

Tweet 1:   *Barbados #Bridgetown JAMAICA, Two cars set ablaze: SANTA CRUZ, Head of the St Elizabeth Police Superintende...*

Tweet 2:   *Ablaze for you Lord :D*

Tweet 1 is classified as a disaster and Tweet 2 is not though the word **ablaze** is used in both the tweets.

1.2   Kaggle Data

This specific problem statement is an ongoing Kaggle competition

https://www.kaggle.com/c/nlp-getting-started/overview/

The data provided for this competition contains real tweets that are filtered using a set of keywords and are manually sorted to identify the real disasters of them. The location data for the tweet posted is also available.

1.3   Solution and Outcome

NLP makes it possible for an artificial intelligence program to receive conversational text, break syntax down to grasp the text's meaning, determine relevant action and respond in a casual manner. By using NLP, we can understand the context of the sentence and understand the real usage of word and thus classify the disaster tweets out of all the other tweets.