

IC23042 Ivan Thomas, Lyndon Taylor, Noah Moufarrij

Data Analytics Challenge 2: CrisisFACTS

Abstract:

Every year Millions of people are affected by natural disasters; they may experience a loss of property or a loved one, or danger to themselves. At these challenging times many people turn to social media platforms to communicate, whether to give information about themselves or a loved one, to get information for rescue and resource or to just vent about their situation. The information in these posts can be very useful to decision makers, rescue workers and volunteers to render aid and scope out the true scale of the disaster. For this data to be useful it is important to verify the accuracy of these posts. Our goal for this challenge was to see to the accuracy of these posts by contrasting the information in the post with the amount of emotion in the posts. This was accomplished by using a *Sentiment Intensity Analyzer* through the *Natural Language Toolkit* in Python. After extracting the emotional data from each post we are able to see which social media platforms are the most informative to any interested parties. We defined this by an overall emotional sentiment close to 0 and a neutral sentiment close to 1 on a scale of -1 to 1. Additionally we viewed how the emotional sentiment expressed changed over time during the disasters, to see when would be the best time to collect data or if the sentiment change wasn't enough to justify delaying data collection.