

# Disaster Tweets Text Classification

Nicole Michaud



# Business Problem

- Agency wishes to monitor twitter in the event of an emergency to *respond faster* and *minimize casualties and/or damage*
- Goal: Create model to analyze tweets to determine whether or not they are about a **real disaster** or not

# The Tweets Data

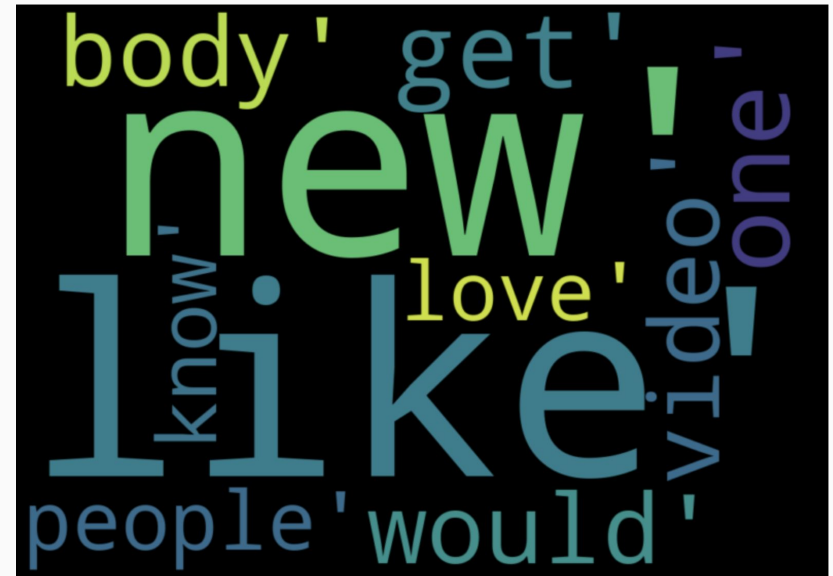
- **7,613** tweets in training set
  - Text
  - Keyword
  - Location
- Probability of a tweet being about a disaster: **43%**
- Probability of not being a disaster: **57%**

# The Tweets Data: Common Words

Top 10 words in **disaster** tweets:



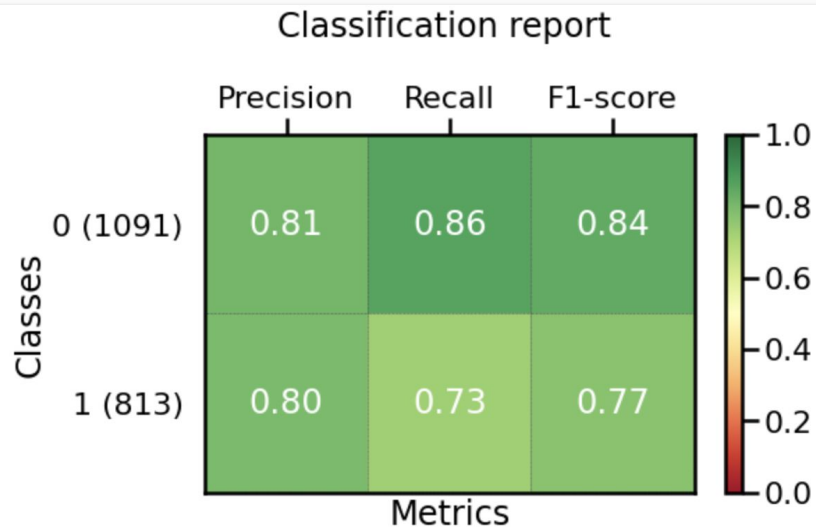
Top 10 words in **non-disaster** tweets:



# Methods

- Models were created and tuned to see which one performed best at correctly determining if a tweet is about a **real disaster** or not
- Evaluated based on **F1 score**
  - compares **false negatives** and **false positives**

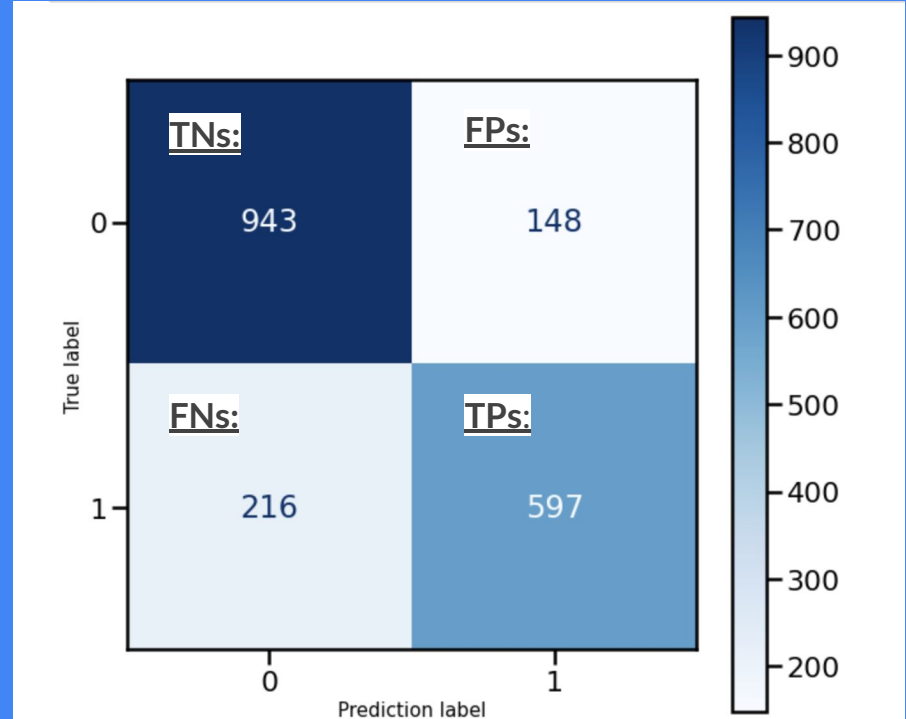
# Results



- Average F1 score: 80.23%

# Results

- **216** False Negatives
- **148** False Positives
  - Both are costly



# Recommendations & Next Steps

- Model should be **deployed** to monitor twitter for disaster tweets
- To try and get an improved model for future use:
  - Try other classifiers/models
  - More tweets data



# Thank you!

- **Any questions?**
- **Contact:**
  - [www.linkedin.com/in/nicole-michaud2](https://www.linkedin.com/in/nicole-michaud2)

