

# THAT TWEET WAS A DISASTER

A KAGGLE COMPETITION:  
REAL OR NOT? NLP WITH DISASTER TWEETS

PRESENTED BY  
DAVID WALKUP



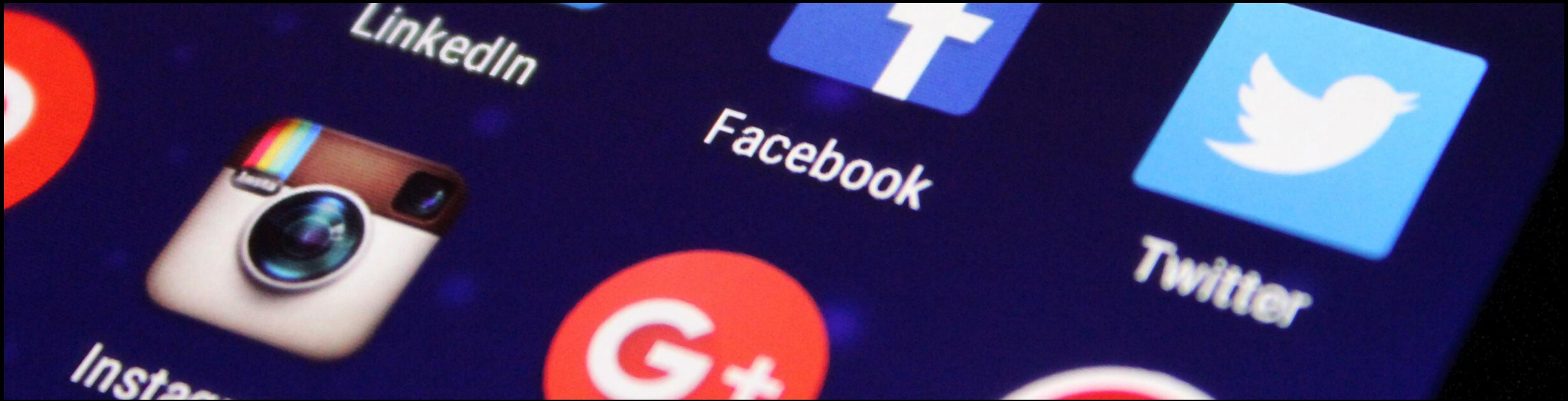
# PRESENTATION AGENDA

- Competition Overview
- Challenges
- Results
- Next Steps
- Q & A

# COMPETITION OVERVIEW

With smartphones everywhere and Twitter accounts commonplace, Twitter has become an important communication channel during emergencies.





People often tweet about disasters and emergencies in real time. This has lead to interest from news and disaster relief agencies in programmatically monitoring Twitter for such announcements.



## COMPETITION GOAL

Given a set of 10,000 hand-classified tweets, train a machine learning model to predict if a tweet is about a real disaster or not.



**THERE WERE CHALLENGES**

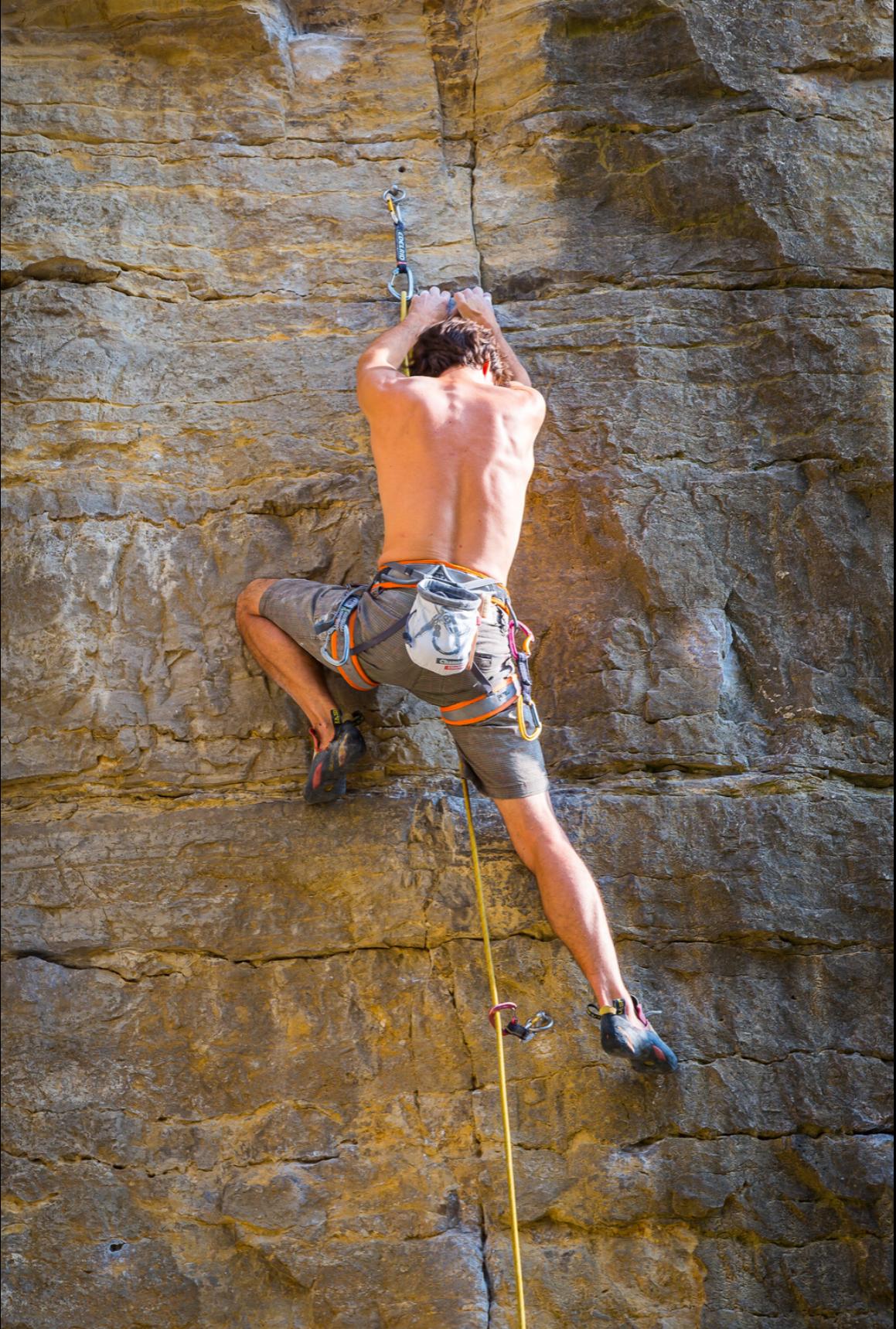
# THE MAIN CHALLENGE

## Text Processing

Non-disaster examples:

"E-Hutch is da bomb ?? <http://t.co/aqmpxz03V1>"

"Evacuation drill at work. The fire doors  
wouldn't open so i got to smash the emergency  
release glass #feelingmanly"



# THE MAIN CHALLENGE

## Text Processing

Disaster examples:

"@moscow\_ghost @sayed\_ridha @Amin\_Akh  
congratulations on capturing a besieged city  
after 3 months of indiscriminate bombing by  
land & air"

"Arsonist Sets NYC Vegetarian Restaurant on  
Fire: Police #NewYork - <http://t.co/Nr7usT3uh8>"



# THE MAIN CHALLENGE

## Text Processing

- Abbreviations and Slang
- Hyperlinks
- Hashtags and Other Punctuation
- Stopwords (i.e., "a," "the," "an")
- Misspellings and Typos

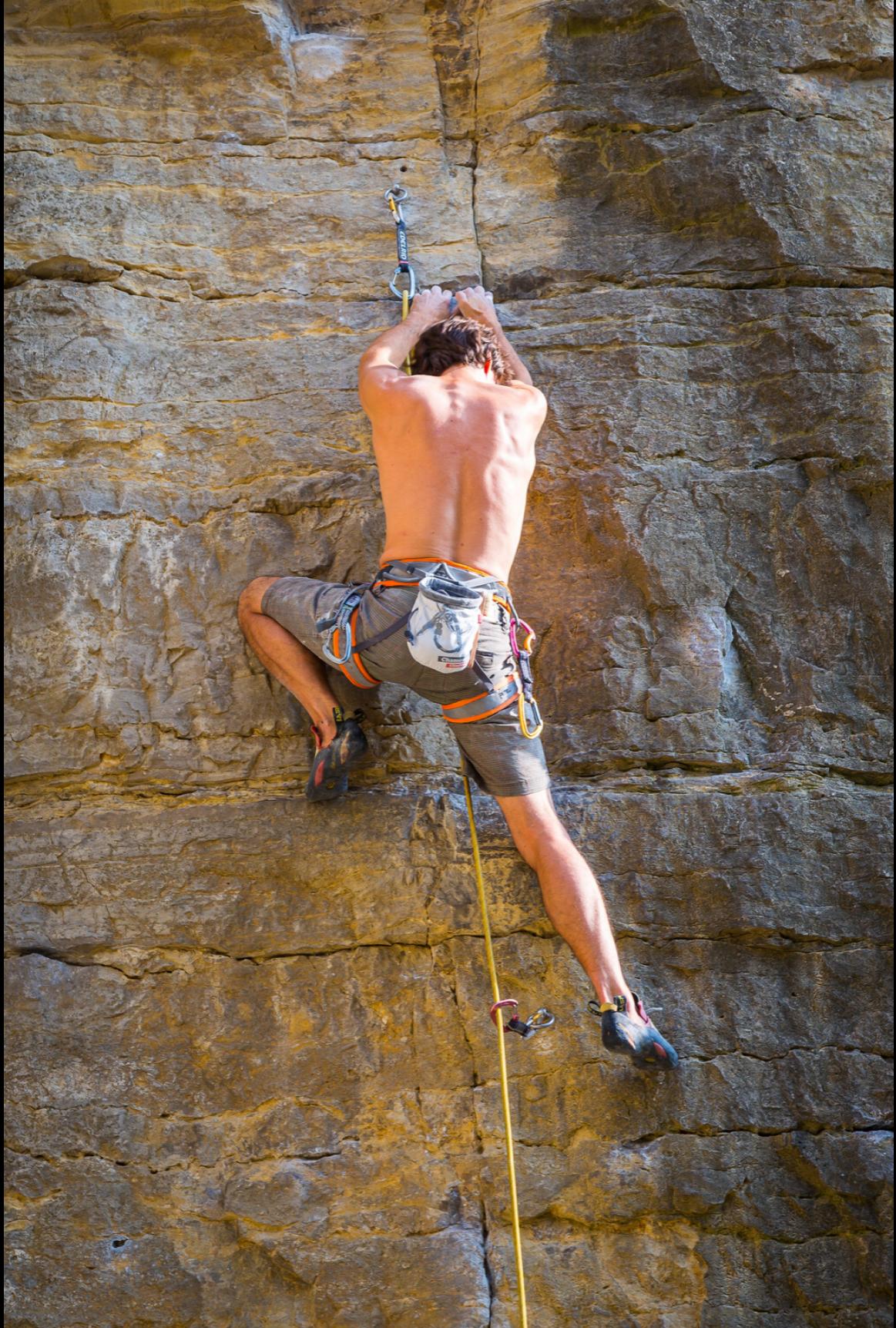


# THE MINOR CHALLENGE

Inconsistent classification of the  
training data

## Example:

"Man Currensy really be talkin that  
talk... I'd be more devastated if he  
had a ghostwriter than anybody  
else...."





# RESULTS

80.78% accuracy in predicting if a tweet was about an emergency or disaster.

## NEXT STEPS

- Testing additional models
- Further hyperparameter tuning



# QUESTIONS?

