

# ML Code Lab

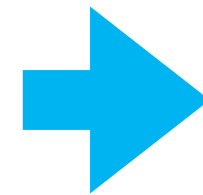
Adam Yala, Regina Barzilay



# Twitter Sentiment Analysis

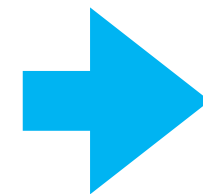


The sun is finally out! #boston



1

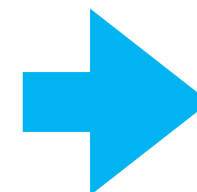
I'm melting in this heat! #boston



0



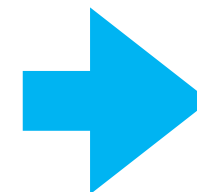
The sun is finally out! #boston



1

$[0, 1 \dots, 1]$

I'm melting in this heat! #boston



0

$[1, 0 \dots, 0]$

# Plan

- Building a classifier from end-to-end
  - Feature engineering
  - Analysis
  - Trying different models
- Next
  - Deep learning approach

# Key Packages

- Numpy
- Pickle
- Sklearn
- scipy
- matplotlib

# Key Resources

- Feature Engineering
  - CountVectorizer- Bag of words, n-grams, etc.
  - DictVectorizer- All other features.
  - hstack - combining feature matrixes

# Code

[https://github.com/yala/  
MLCodeLab](https://github.com/yala/MLCodeLab)



# Questions?