

# Machine Learning Design Project

Annotation

# Design Project

- Data prep
- Train model to predict outputs

# Data Preparation

- Annotate assigned data from language survey
  - [https://docs.google.com/forms/d/e/1FAIpQLScLlasz4W-ci4TZLJT2\\_PINhMhPbhhaBnyZZhTnc5VoejYNwQ/viewform](https://docs.google.com/forms/d/e/1FAIpQLScLlasz4W-ci4TZLJT2_PINhMhPbhhaBnyZZhTnc5VoejYNwQ/viewform)
- Mark named entities on Twitter data

# Named Entities

- Name of person, organization, location, product

**Tesla** chief **Elon Musk** did not like **Singapore** when the government taxed its **Model S** car.

# Named Entities

- Annotate the outermost entity

✓ I went to the **Bank of China**

✗ I went to the Bank of **China**

# Named Entities

- Named entities:
  - Hashtags are names only if they are part of the sentence
    - The Final Episode Of **#TheBachelor** Draws Season Long Fans
  - @-mentions only if they are not usernames
    - round 2 (@ **Club Tryst** in **Singapore** w/ @coldsteel\_lex)
- Not named entities:
  - Dates and times ("Tuesday"), seasons ("autumn"), holidays ("Deepavali"), addresses, web trends (#ThrowbackThursday)

# Sentiment Polarity

- Is the sentiment on the entity positive / neutral / negative?

*neutral*

*neutral*

*negative*

**Tesla** chief **Elon Musk** did not like **Singapore**

*positive*

**Singapore** is great !

# Demo



# Design Project

- Data & detailed instructions will be released after survey results & assignment
- Graded on completeness & sample accuracy
- Annotation is the easy part :)