

# Tweets Sentiment Analysis

Kuo-Wei Ho<sup>1</sup>, Hao-Chien Wang<sup>2</sup>

<sup>1</sup>NTUEE

<sup>2</sup>NTUPhys

Data Science Programing,  
July 2019

# Table of Contents

# Problem

Given a set of data containing 1,600,000 tweets and the sentiment of each tweets. Create a model that can analyze sentiment of new tweets.

Data: <https://www.kaggle.com/kazanova/sentiment140>

Table: Data example

sentiment	Post ID	User ID	tweets
0	1467814192	Ljelli3166	blagh class at 8 tomorrow
0	1467821455	CiaraRenee	I need a hug
4	1677796507	FoodAllergyBuzz	@otibm! Thx for the tweet!
4	1677796519	lakido	Sunshine.....! LOVE this weather!!!

0: Negative

4: Positive

- *GloVe: Global Vectors for Word Representation* by Stanford University.
- *RNN: Recurrent Neural Network*

- 1 Clean the data: remove non-UTF8 symbols, numbers and urls.
- 2 Combine all tweets into one string and tokenize.
- 3 Feed the tokens to GloVe to generate word vectors.
- 4 tokenize all tweets and search each words in the vectors to transform it into a list of matrices.
- 5 Train RNN with the list of word vectors.