COMP5541 Machine Learning and Data Analytics

# **Due date**: 24 November (23:59)

**Stock Market Prediction with News**

# **Project Objective**

Stock market prediction is a growing hot topic in artificial intelligence. Many financial companies have developed their own invest robots. In this group project, you are required to use 2-year stock market movement and daily news (tweets) to predict the stock price. You are expected to understand the pipeline of machine learning, including feature extraction, model selection and parameter optimization. Moreover, you can develop a practical invest robot by yourself.

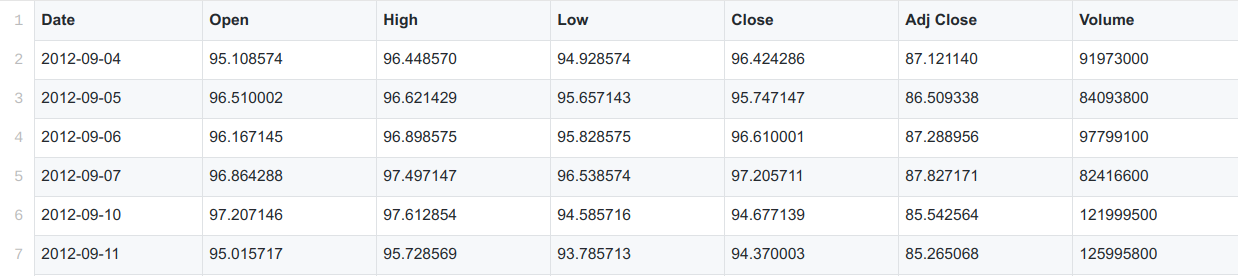
Attention: you must write your own **Python** program.

# **Task Description**

Input: History of stock market movement, corresponding tweets each day and the tweet data of the next seven days.

Output: The stock market adjusted close prices (‘adj close’) of the next seven days

Train/Test separation: 2014-01-01 to 2015-12-20 for training, 2015-12-21 to 2015-12-31 for evaluation.





Sample Data

Hint: You can use [pandas](https://pandas.pydata.org/) to load the dataset

# **Project Requirements**

Given eight stocks, you are required to use the history of stock market movement and some tweets (text data) to predict the adjusted close (‘adj close’) stock prices of the following seven days.

Basic Requirements

1. Build a stock price prediction system which can predict the adjusted close stock price based on the history of stock market movement and some text information.
2. No hard-coding in this project

Advanced Requirements

1. Try to develop a deep learning model (e.g., Recurrent Neural Network)
2. Try to visualize the result and model

# **What to Hand In**

1. The Source Code of Your Program

* File submission requirement goes here
  + A readme file describes the structure of your program and how to run it
  + Add annotations in the front of each Python file to specify its usage

1. Written Report

* Describe the architecture of your system, including the input features, the model designed, the loss function, the optimization method, e.t.c..
* The functions of your system and the tools or approaches you use.
* The contributions to this system of your each team member
* Any other advantages of your system that you want to highlight

Submit your report and source code to Blackboard.

# **Evaluation and grading scheme**

1. We use mean square loss (MSE) to evaluate your predictions.

2. Your grade is determined by the evaluation results and your report.