

# Homework #4

## CS 525/DS 595, Spring 2020

---

100 points total [6% of your final grade]

**Due:** April 17, 2020 by 11:59pm

[no submission will be accepted after April 20, 2020 at 11:59pm]

**Delivery:** Submit via Canvas

---

### Donald Trump vs. Joe Biden

In this homework, you are going to learn the basics of the social media mining by collecting Twitter data, preprocessing the data, and conducting exploratory analysis.

#### Task 0: Reading Chapters 1 and 9 of [Mining the Social Web](#), edition 3

Search “Mining the Social Web” in WPI library. You can access the online version for free.

The chapters describe how to access and use Twitter APIs. You may refer to [sample code](#) in the book:

#### Task 1: Collecting Twitter Data

In this task, you will collect some Twitter data by using [Twitter APIs](#), especially, using Rest APIs.

By using a search API, collect at least 1,000 tweets containing a keyword “donald trump”, and another 1,000 tweets containing “joe biden”. Store collected data in the JSON format. Add the following conditions when you use the search API:

- Tweets should be written in English.
- Remove/filter retweets to get more meaningful information if you can get at least 1,000 tweets.

*Include a few sample tweets in your report. Also, write the detailed steps that you followed in order to collect your Twitter data.*

#### Task 2: Exploratory Analysis

In order to gain clear understanding of the data that you preprocessed, you will conduct exploratory analysis. You may use external libraries (e.g., scikit-learn, nltk) for the analysis.

- A time series figure with the number of tweets per day over time for both candidates.
  - Perform two more interesting analysis of your choice (e.g., sentiment analysis, clustering and so on).
  - Based on the above analysis, who will be elected in 2020 presidential election? (just assuming that Trump and Biden will be the final candidate from each party)
  - *Report what you analyzed, found and learned.*
- 

What to turn in:

- Submit to Canvas your report (a PDF) and your code.
- This is an individual assignment, but you may discuss general strategies and approaches with other members of the class (refer to the syllabus for details of the homework collaboration policy). At the top of your report, please write the names of classmates you consulted and the nature of your discussion.