Final Project Assignment (Group): Informatics on Election Integrity Twitter Dataset

Points: 250 points.   Assignment Date: October 08, 2019   Due Date: December 09, 2019

Objectives

1. To define and describe a problem in informatics focused on the Election Integrity Twitter Dataset
2. To design and develop your solution to address the problem
3. To carry out the various steps of Informatics to solve the problem: data preparation, data cleaning, data pre-processing, data analysis, and data visualization
4. To implement programs in Python to perform the above steps of informatics on the data downloaded
5. To write a report to document your problem and solution, your informatics steps, outcomes and results of your informatics work, insights and lessons learned, and potential future work
6. To present your final project well using slides and a software demonstration
7. To learn about teamwork and work as a team

Dataset

Referring back to the lecture by Dr. Deepti Joshi of the Citadel on September 08, 2019, we see that there are numerous countries included in the Election Integrity Twitter Dataset.   There are multiple folders for each country's dataset: account information, tweet information, media, etc. Explore these folders and datasets. These folders have files that are connected through tweet IDs and user IDs, allowing one to analyze multiple tweets by the same user, as well as tweets under the same hashtag, for different regions and different time periods.

Final Project Problem

For this final project, identify a specific problem in Informatics that your team will investigate given the datasets. For example, are there interesting questions to answer? Are there patterns that you want to find? Are there trends to be observed? Explore the datasets to get yourself familiarized with the data first, and then form your questions that will drive your Informatics efforts.

The tweets have both content (the texts or media or links inside a tweet) and also metadata (e.g., user, time stamp, location, etc.) Analysis could be done to just the content, or the metadata, or both. Your informatics problem will generally be more meaningful if it involves both content and metadata in your investigation and analysis.

Requirements and Handin

1. The submission deadline for all handins is December 09, 2019, 11:00 AM, which is one day before our Demo Day. **Late handins will *not* be accepted or graded.**

2. **Final Project Proposal Q&A** (0 points but REQUIRED):

   a. **October 31, 2019:** Each team is required to turn in a 3-page Final Project Proposal and present their proposal (using Powerpoint slides) in class and participate in a Q&A discussion.

   b. **The Proposal must include at least the following:** (1) a Team Name, (2) The names of all team members, (3) Proposed informatics problem, (4) Proposed tasks to solve the informatics problem, and (5) Specific tasks and responsibilities for each team member.

3. **Final Project Report** (100 points): Each team is required to handin a final project report. The following sections are required:

   a. **Introduction** (15 points): This section describes your informatics problem

   b. **Data Preparation** (15 points): This section describes your data exploration strategy and how you extract with the data that you used for your final project.

   c. **Data Cleaning & Pre-Processing** (10 points): This section describes your data cleaning and pre-processing strategies used. Must refer to any Python programs that you used to accomplish this step.

   d. **Data Analysis** (20 points): This section describes your data analysis solution. Must justify your analysis and findings. Must refer to any Python programs that you used to accomplish this step.

   e. **Data Visualization** (20 points): This section describes your data visualization strategies. Must justify your strategies. Must refer to any Python program that you used to accomplish this step. Must also provide the infographic.

   f. **Conclusions** (10 points): This section documents any insights or lessons learned from this Informatics assignment

   g. **Appendix** (10 points):

      i. This section describes your overall approach to implement the Python programs and the list of all Python programs that you implemented and their purpose.

      ii. This section must also include the 3-page proposal.

4. **Data and Programs** (50 points): For each team:

   a. You are required to handin a screen capture of your "testing session" using your programs. (10 points)

   b. You are required to handin all program files. (10 points)

   c. You are required to handin all input and output files. (5 points)

5. **Final Project Demo** (50 points): For each team:

   a. You are required to present your Final Project using Powerpoint slides and execution of your programs at real-time on Demo Day (**December 10 and 12, 2019**)

b. Your presentation should cover all sections of your Final Project report adequately (35 points)

c. Your demonstration of your programs should proceed smoothly showing how you obtain the output from running your programs (15 points)

6. You are required to handin online the above using **http://cse.unl.edu/handin/**