

Final Project Milestone 2

1 **What is your group #?**

Group 4

2 **Have you met/communicated with your fellow teammates?**

Yes. We have communicated through email and the Group Collaboration space within Canvas.

3 **Have you met/communicated with your assigned TF? If not, please provide a reason.**

Yes. We have asked him a few questions regarding project scope via Canvas.

4 **Has your team formulated a well-defined question to address in your project, based on the project description and references? If so, please write down the question. If your team hasn't done this yet, that's okay!**

Our main question is the following:

For users tweeting a trending hashtag (we will choose one hashtag), can we differentiate between bots and humans by NLP alone by observing the body of tweets for those users via Natural Language Processing (topic modeling or sentiment analysis)? Classification? Combination of both?

5 **Briefly describe your team's plans for work to be completed by Nov 28 (milestone three). Please assign specific tasks to team members and deadlines for when these tasks are to be completed.**

- Complete Data Engineering and EDA by Nov 21. Kiet and Shaswati will complete Twitter data collection, cleaning, and feature engineering by Nov 21. Mary will create a function to enrich the Twitter dataset with Indiana University's Twitter Bot-o-Meter.
- First Data Pull due Nov 8 (Kiet and Shaswati for Twitter Data, Mary for Twitter Bot-o-Meter)
- EDA 1, Nov 8 -14 (Split up features among us)
 - We will need to decide how to deal with missing values
 - Identify how each feature should be classified (categorical, numeric, dummy encoded, etc)
 - Identify features that need to be normalized
- Make changes to Dataset based on EDA 1 results (Nov 15-21) Kiet and Shaswati
- EDA 2, Nov 21-23 Split up? Mary can start building functions to perform EDA while the dataset undergoes final grooming.
 - Visualize features to identify Collinearity
 - Visualize features to identify predictive power (histograms of feature values vs. response)

- How do the feature behaviors change if they are plotted hourly?
Daily?
- Each person will document their contribution (description of data or EDA findings) by Nov 26. -- The group will create a template for the type of information we want to document for each.
- Mary will write the final version of milestone 3 and submit to Canvas on Nov 28.