Sentiment Analysis of Fake/Real News Case Study Rubric

Due: TBD

Submission format: Submit GitHub Repository Link to Canvas

General Description: Submit to Canvas a link to your GitHub repository for this case study

Why am I doing this? The goal of this project is to give you an opportunity to demonstrate your technical and analytical skills through a meaningful application. You will conduct sentiment analysis on fake and real news articles surrounding the 2016 presidential election to explore whether sentiment plays a role in distinguishing between fake and real news, connecting your work to a real-world issue in public governance.

What am I going to do? You will provide a deliverable that covers all the requirements of this case study. You can find instructions and resources for the study here: https://github.com/kyletran511/CS-Sentiment-Analysis-of-Fake-Real-News

In this case study, you will conduct sentiment analysis on both fake and real new articles published in proximity to the 2016 presidential election. You will propose a hypothesis, perform exploratory analysis on the data, and a form of statistical analysis to support your findings.

The final deliverables should include:

- A GitHub repository containing: your code, data, visualizations produced, data appendix, and detailed README
- In your analysis, you should:
    - Perform some form of sentiment analysis (i.e. using VADER python package)
    - Perform some sort of statistical analysis (i.e. t-test) to test your hypothesis
    - Generate visualizations to support your EDA
- You have the freedom to decide what types of technologies you would like to use for your analysis

How will I know I have Succeeded? You will meet expectations on this case study when you follow the criteria in the rubric below.

Formatting	• One Github Repository (submitted via link on canvas)
	• To ensure **reproducibility**, the repository will adapt parts of the TIER Protocol 4.0. In a nutshell, the top level page of the repository should contain:
	  ○ A README.md file (which auto displays)
	  ○ A SCRIPTS folder
	  ○ A DATA folder
	  ○ AN OUTPUT folder

README.md	• Goal: This file serves as a guide to help users navigate the project you have produced from this case study. • Make an H2 (##) section explaining the contents of the repository

	● Section 1: Software and platform section (Up to Your Choosing) ○ The type(s) of software you used for the project. ○ The platform (e.g., Windows, Mac, or Linux) you used. ● Section 2: A Map of your documentation. In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder. ● Section 3: Instructions for reproducing your results. In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise
SCRIPTS folder	● Goal: This folder contains all the source code for your project. Throughout all your scripts, you should include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is. ● You code must include some form of sentiment analysis, statistical component, and 3 visualizations
DATA folder	● Goal: This folder contains all of the data for this project. ● You should include the initial data, and the final data analyzed. o ***You are expected to have a different final data analyzed than the initial data provided*** ● A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics. This file should be organized in sections, with a section for **each** dataset analyzed. More information: https://www.projecttier.org/tier-protocol/protocol-4-0/root/data/analysisdata/data-appendixfile
OUTPUT folder	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc. ● Importantly, any information like tables, figures shown in your presentation should be here. ● **Use informative names for your files**.
References	● All references should be listed at the end of the document ● Use IEEE Documentation style (link)

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for providing a template for this rubric. This structure is pulled from Streifer & Palmer (2020).