Case Study: LinkedIn Engagement Model

Due: TBD
Submission format: Upload link to github repo to canvas
Individual Assignment

General Description: Using Natural Language Processing (NLP), generate an analysis on your LinkedIn profile.

Preparatory Assignments: All previous Data Science curriculum and practice.

Why am I doing this? As most people intend to leave college with a job, and as LinkedIn has become a popular mode for people to network and build a professional online presence, your deliverable will help provide measures of your LinkedIn profile's engagement and performance. You can then use these metrics as a basis for actionable insights to grow your presence as a professional and aid in your job searches. Based on previous research on the LinkedIn Job Description data set, an analysis provided some basic LinkedIn profile formatting advice that would help a profile stand out in the LinkedIn algorithm. Now, it is your turn to apply this analysis to specifically your presence on LinkedIn.

> **Learning Objective:** Practical applications of data science basics like random forests, and building familiarity with NLP models

What am I going to do? First, read the provided materials in this folder, starting with "Analyzing Linked In Content: A Practical Application of Data Science." Then, taking inspiration and recommendations from previous analyses on LinkedIn Job Descriptions, analyze your own LinkedIn profile and presence. Using a Natural Language Processing model like those in spaCy and a random forest model, list three to five ways you can improve your very own LinkedIn engagement metrics and what analytic outcomes you based those recommendations on. Deliverables include:

- A github with:
    - A dataset of your LinkedIn profile metrics (views, likes, comments, engagement score)
    - A README.md
    - Code of your exploratory data analysis
    - Code of your two models
        - A spaCy Natural Language Processing Model
        - A random forest
    - A presentation of actionable insights to improve your LinkedIn engagement

Tips for success:

- Be honest with your performance. You will get more out of this assignment if you view your engagement from an objective perspective.
- Explore spaCy's capabilities. It is a powerful tool and, among NLPs, easy to use.

How will I know I have succeeded? You will meet expectations when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	● One Github Repository ● The top level page should contain ○ A README.md file (which auto displays) ○ A LICENSE.md file (use MIT as default) ○ A SRC folder ○ A DATA folder ○ A FIGURES folder
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ● Use markdown headers to divide content ● Make an H2 (##) section explaining the contents of the repository ● SRC section ○ Make an H3 section for installing/building your code ○ Make an H3 section for usage of your code ● DATA section ○ (This one is tricky. Your data may (or not) fit in repo) ○ Data Dictionary (use markdown table formatting) ○ Data Files or Link to data if it doesn't fit on github ○ Relevant notes about use of data ● FIGURES section ○ Table of contents describing all figures produced and summarizing their takeaways ○ Use markdown table formatting · ● REFERENCES section ○ All references should be listed at the end of the Readme.md file (Use IEEE Documentation style (link)) ○ Include any acknowledgements
Figures folder	● Goal: This folder contains all of the figures generated by your project. ● Any figures you use in your presentation should be placed here. ● Include with every figure relevant notes about the figure. ● Any figures, plots, or graphs you generated with your code should

	be placed here.
DATA Folder	● Goal: This folder contains all of the data for this project. ● If your data fits in the GitHub, place all of it here. ● If your data does not fit in the GitHub, use a single file explaining the process to obtain the dataset. ● This should include your LinkedIn engagement metrics as provided by LinkedIn, which includes views, likes, comments, and engagement score.
SRC folder	● Goal: This folder contains all the source code for your project. ● Include all code files you produce. ● Include supplemental documentation as necessary, especially if it is too detailed/verbose for the overall readme. ● The accuracy of your model is not of importance. Focus on practical usage of data science tools and techniques. ● Your code should include your exploratory data analysis. ● Your code should include an implementation of a spaCy NLP model. ● Your code should include an implementation of a random forest model.
License	● Goal: This file explains to a visitor the terms under which they may use and cite your repository. ● Select an appropriate license from the GitHub options list on repository creation. ● Usually, the MIT license is appropriate.
Presentation	● Goal: Contextualize and explain the project prompt to an audience. Then, explain your results from conducting your analysis. ● This presentation is focused on the big picture, especially the results of your analysis. Why do you think an NLP/random forest model was chosen for you? What was your output? How do you plan on using the output you generated?