CS - Twitter Airline Sentiment Case Study

DS 4002 Fall 2023- Siona Pathak
Due: May 2024
Submission format: GitHub repository and Powerpoint (submitted by link to canvas)

Individual Assignment

General Description: Submit to Canvas a link to your Github Repository and a PPT of your presentation.

Why am I doing this? Case studies are a great way to apply technical and theoretical knowledge to real world settings and applications. Case studies provide students with opportunities to demonstrate problem-solving abilities. You will learn to design methodologies and apply relevant data science techniques to reach conclusions. Additionally, real world data is not controlled and cleaned for you like in a classroom environment. It is messy and in high volumes, a fact that poses a challenge for you as a student in regards to adding a step of data cleaning. You will get to analyze real data, in this case tweets, to derive insights for companies to capture value. The experience of getting your hands dirty with real world data to find actionable solutions to problems is invaluable and will serve you well in a job or research setting.

- Course Learning Objective: Apply theoretical data science knowledge to create functional results
- Course Learning Objectives: Apply critical thinking to adapt model to real world situation
- Course Learning Objectives: Compile findings to find conclusions and present findings

What am I going to do? You will first utilize sentiment analysis techniques to determine which keywords appear most frequently in airline tweets. You will utilize pre-trained APIs to execute this task. Then, you will combine the code and results of this in a Github repository that should have a level of detail and organization that make it easy for anyone to read and understand. Finally, you will compile your findings into a presentation for your peers to understand their real-world impacts on your job and the industry as a whole.

Tips for success:

- Understand the techniques you use - the model is pre-trained but you should be able to adapt it to fit your needs!
- Stay focused - Allot ample time for this project and focus your full efforts on it as you work.
- Be determined - this may not come easy but nothing good ever does! Keep trying and don't give up.

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

Spec Category	Spec Details
Github Repository	- One Github Repository (submitted via link on Canvas) - Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. - The top level page should contain - A README.md file (which auto displays) and contains: - SRC section: Make subsections for - - Installing/Building your code - Usage of your code - DATA section - Data Dictionary (use markdown table formatting) - Data Files or Link to data (given) - Relevant comments on data - FIGURES section - Relevant comments on all figures - REFERENCES section - All references should be listed at the end of the Readme.md file (Use IEEE Documentation style (link) - A LICENSE.md file (use MIT as default) - Goal: This file explains to a visitor the terms under which they may use and cite your repository. - A SRC folder - Goal: This folder contains all the source code for your project. - You will be given some foundational code with the sentiment analysis model to work off of - include the additional code you produce as part of this file - A FIGURES folder - Goal: This folder contains all of the figures generated by your project - Include with every figure relevant notes about the figure

	■ If you are going to use a figure in your presentation place it here
	  ○ References
	    ■ All references should be listed at the end of the document
	    ■ Use IEEE Documentation style ([link](#))

Presentation	● About 7 slides
	● PDF format for submission to collab
	● Generate the slides through the program of your choice
	● Slide numbers (except for title slide)
	● Order
	  ○ Title & Outline
	  ○ Motivation/Context/Hypothesis/Research
	  ○ Question/Modeling Approach/Goal/Etc.
	  ○ Data Explanation/Acquisition
	  ○ Analysis Plan and Justification
	  ○ Tricky Analysis Decision
	  ○ Bias and Uncertainty Validation
	  ○ Results/Conclusions
	  ○ Next Steps
	  ○ References/Resources/Acknowledgements
	  ○ Closing Slide
	● General note: Each section can take as many slides as it needs to unless specifically indicated as 1 slide.
	● General note: Do not take the exact slide count as critical. When you practice you can time out the talk to hit the 7-minute mark. Some slides need more, and some slides need less time.