Final Capstone Product

Build a Model that can be considered a “Data Science Product”.

It should solve a real problem or create something of value.

Deliverables:

* A one-page proposal to be approved by the data science team here, before you begin
* A 20 Minute Presentation on your product that argues not only for its relevance, but why this implementation is best, and the technical tools you will use to get there. This presentation should include a demo!
* A Jupiter notebook that builds the model and shows your machine learning tuning process
* Defend your model in a 60-minute mock interview (the interviewer will attend your presentation and read through your notebook beforehand)
* A final 60-minute mock interview on data science generally

Data Science Requirements:

* Utilize at least one specialization in your pipeline
  + Advanced Natural Language Processing and Neural Networks
* Include supervised and unsupervised techniques

Guidelines:

* Your data set will be hugely important here. Though we've relied on pre-aggregated data before, for this assignment gathering your own data through scraping or other techniques may allow you to make a more impactful or novel product. It will likely need to see a continuous flow of data in order to be relevant as a data solution.
* Emphasize why your implementation is the best implementation, not just an implementation
* Remember, a good product is more than just metrics. It solves a problem.
* You do not have to make a front end or a user interface for this product

**Processing Complaints**

* Purpose: to process complaints/ concerns/ questions relating to product or business and address them more efficiently.
* Scrape information on reviews, complaints, and questions (questions will probably be processed differently than complaints a reviews).
* Use NLP and Neural Net algorithms to process the text data and classify them into topics
* Within each topic, use NLP and Neural Net Algorithms to decide which complaints, questions are most important to address (multiple similar complaints, common question, etc.)
* Return ranked result of these concerns in each topic

**Market Information Extraction**

* Purpose: to extract information from public news and social media sources that can affect one’s company’s market. This may direct or indirect effects i.e. competition information, market trends, consumer outlooks and concerns, new developments, new technologies, etc.
* Scrape textual information from the web on news relevant to your companies’ market. This information should be separated by topics such as consumer information, competition information, market trends and new technologies.
* Within each topic, the data can be populated by pertinence (hot topic or not) into a dataset which companies can view and then process accordingly.
  + Popular Concerns for companies include:
    - Environmental awareness as it affects consumers
    - Social media as it affects consumers

**Above Ideas could be Connected:**

* Consumer Reports on multiple companies to be used by both companies and users.

**Scraping program: Closure**

**Proposal**

The proposal should be approximately one page long and answer the following questions:

* What is the problem you are attempting to solve?

Tech companies want to assess their products’ performance. One way of doing that is analyzing consumer reviews, feedback, thoughts and assessments on their products.

Most companies have their own forum and methods of surveying for collecting feedback and reviews, but most people don’t place general reviews and opinions directly on those forums unless they feel strongly about something specific that’s affecting them, so there’s a bit of a selection bias.

* How is your solution valuable?

This project will aggregate information consumer reaction and opinion on the product from popular mediums like the news and social media. The company can then use this information to learn about the biggest talking points, that people are saying about their product. My product will also classify the sentiments of the reviews. The company can use this product to optimize their product for future version releases.

* What is your data source and how will you access it?

I want to scrape textual data from TechRadar.com concerning one product, identified by a key word.

* What techniques from the course do you anticipate using?
  + Natural Language Processing
  + Neural Network
* What do you anticipate being the biggest challenge you’ll face?

My biggest Challenge will be scraping the data and cleaning it into a form I can process and explore.