Task List

CS_122_Quarter_Project

Jasmina Scekic, Will Fisher, Raphie Rosen, and Gloria Huang

Project Description

Many food options that are advertised as "healthy," "natural," or "diet-friendly" actually contain a number of "hidden" ingredients that actually are not good for you. The goal of this project is to create a program to evaluate food options based on the quantity and type of artificial and unhealthy ingredients they contain (e.g. sugars, low-calorie sweeteners, artificial colors, etc).

We will gather a list of the ingredients we want to look for and search the ingredient lists of products at several large online grocery stores. Each product will be assigned a health score based on the ratio of unhealthy to total ingredients and possibly also based on the quantity of the unhealthy ingredients relative to the rest of the ingredients. The relative quantity would be determined by the position of an ingredient in the list, since ingredients are listed in order of quantity. For example, a protein bar with sugar as the first ingredient would score lower than if the bar had sugar as the final ingredient.

The user will be able to select a category or brand to explore and determine the healthiest option available based on our health score. In addition, the user will be able to explore specific products and both view the health score and see how it was calculated (based on number and type of ingredients in that product). This project will allow the user to make a more informed decision about purchasing foods based on their ingredients, many of which are unrecognizable to most people.

Data Sources

- Target:
  - https://www.target.com/c/chips-snacks-cookies-grocery/-/N-5xsy9?intent_filters=
- Specialty online health food stores (i.e. Vitacost, Direct Eats)
  - https://www.vitacost.com/snacks-7#
  - https://directeats.com/food/snacks?sort=Realware.Plugin.ProductSort.PriceLowHigh
- FDA Additives List:

- ○ https://www.fda.gov/food/food-additives-petitions/food-additive-status-list
  - ○ https://www.cfsanappsexternal.fda.gov/scripts/fdcc/index.cfm?set=FoodSubstances
- ● Additives to Avoid:
  - ○ https://mphprogramslist.com/50-jawdroppingly-toxic-food-additives-to-avoid/

Task List

1. Scrape data on granola bars and other snacks marketed to be healthy and their ingredients from Target, Vitacost, Direct Eats, and other online sellers
2. Use FDA and other health advisory sources to determine which ingredients to include on a list of unhealthy ingredients in healthy snacks
3. Use the ingredient list to compute a health score for each snack that we have listed in our database.
4. Make an interface such that health snacks and their scores can be searched for by brand, type of snack, ingredients, score, and other traits.

Timeline

End of week 4: Establish data sources for lists of products and ingredients

End of week 5: Begin scraping data, hopefully get data in a database by end of week

End of week 6: Clean data, deal with None cases/matches when joining data from different sites

End of week 7: Build the model to rate each item

End of week 8: Start to build the application and make it interactive in terminal

End of week 9: Cleaning everything up, making it work, potentially making it run on the web.