

Jordan Deagan

Professor Doucette

Intro to Data Mining & Machine Learning

8 October 2018

Project Proposal

The question I want to ask is “Of beef jerky, trail mix, and protein bars, which is more often purchased, and which is preferred in larger purchases?” What I mean by this is that I want to know which trail snack food is more popular in general, which one is more likely to be bought along with other items, and I want a look at what those other items might be, whether it is other food items, camping stuff, or random products. My desire to ask this is to find out primarily which is popular, and through checking them in large purchases, see if certain buying patterns emerge around buying hiking and camping gear and these items, or if they are primarily bought with other items or individually. I would use a dataset of Amazon purchases to find the information. I would calculate the base percent purchase rate of each item being in a purchase to answer the first half of the question, then using apriori I would run through the purchases, keeping only purchases larger than a specified value that contain one of the three items.

I plan on running the purchases dataset through apriori filtering only for purchases with one of the specified items, calculate percentage of purchases from all of those, and then continue running it to gather a set of purchases containing the item and at least so many other items. I will then run the other two items through in the same manner and compare the results of the initial check and the large size check to see which product is more common, and then I

will run the code a few more times to see what is most oftenly bought with each item.