

MTH/CSE 4224 Intro to Machine Learning Project 2

Project Description

Pursue a machine learning project including use of unsupervised learning. You will propose a problem and submit a proposal outlining the goals of your study, intended data sources, and initial plans for cleaning/preprocessing and models you will use.

Proposal

Goals of study

I would like to look into purchasing demographics and behaviors of shoppers and to utilize clustering to identify distinct purchasing patterns within unique groups that share demographic information in my dataset.

Intended data source(s) and initial plans for cleaning and/or preprocessing the data

I have found a [Customer Segmentation: Clustering](#) dataset on Kaggle that I think would work well for my intended study. This dataset has a lot of information in it that may not be necessarily useful for what I intend to look at. For example, I will most likely remove any data relating to whether or not a customer responded to a specific marketing campaign, if a customer submitted a complaint, cost associated with contacting a customer, revenue associated with a successful campaign, and whether or not an individual responded to any marketing campaign. These features don't necessarily have to do with the demographics or the behaviors of the customers.

Initial plans for methods you will use

I would like to use three different methods for this clustering: K-Means Clustering, Hierarchical Clustering, and DBSCAN.

Your team

I will be working alone for this project.