Instacart Market Basket Analysis

Instacart is a grocery delivery application that works in the following way- users can select products through the app, and then personal shoppers review the order, shop in-store and deliver the products to the users.

We got the data from Instacart's Market Basket Analysis on kaggle.com [here](https://www.kaggle.com/c/instacart-market-basket-analysis/overview).

The Instacart data set is anonymized and contains samples of over 3 million grocery orders from 200,000+ Instacart users. <br>

Business Problem:

On its website Instacart has a recommendation system, that suggests the users some items to buy again. Our

task is to predict which items will be reordered on the next order.

We approach this using 3 methods: -Logistic Regression: to predict product reorder

-Association Rule Mining: utilizing the apriori algorithm and perform market basket analysis to generate association rules with different values of support and confidence, to predict what products are closely associated with or frequently bought with what products

-Gradient Boosting Model: to predict product reorders as well as produce association rules within the test dataset

Persona Panels

<h3>Description</h3>

Persona Panels is an AI Market Research company that provides persona animated bots- designed to mimic target consumer behavior. Every day the bots consume information over the internet that further influences their future choices. These bots are then used to perform market research to test client products and services. We were tasked with analyzing this monitoring data for the (Millennial Persona Panel)[https://personapanels.com/millennial-persona-panel/] to compare and evaluate the bot performance. We were supposed to find interesting similarities, differences and any such valuable insights that might help understand the personas better.

<h3>Process</h3>

Each persona has a set of keywords relating to their interests associated with them; these are monitored daily. I looked at the top most observable and least observable keywords for each persona, common keywords among personas, and the trends. I categorized all keywords with themes like Political, Lifestyle, Family Oriented, and Technology, to see how the interests were distributed among all keywords.

Time series analysis of the keywords was performed using the appropriate time series model to view future upward or downward trends in the interests of the personas.

<h3>Analysis</h3>

Book Recommender System

As one of the world’s most influential reading sites, Goodreads provides a platform for people interested in talking about books. It can be hard to find your favourite next book. A lot of recommender systems exist for video content, and it would be nice to have an effective system for books too!

In this project I built a recommender system based on Netflix’s User- Based Collaborative Filter algorithm to predict book recommendations for over 53,424 users with over 1 million ratings.

The data is from Kaggle’s (goodbooks dataset)[https://www.kaggle.com/zygmunt/goodbooks-10k]. I also compared the results with (recommenderLab’s)[ <https://www.rdocumentation.org/packages/recommenderlab/versions/0.2-6>] UBCF implementation and compared results using different user clusters sizes.