Grocery Recommender - Capstone Two

Problem statement

Gather data on grocery purchases to recommend additional products to consumers using the data science method:

1. Problem Identification
2. Data Wrangling
3. Exploratory Data Analysis
4. Pre-processing and Training
5. Modeling
6. Documentation.

Context

Provide useful product recommendations to help increase sales and improve customer satisfaction with the system.

Criteria for success

- Learn how machine learning algorithms work for content based, collaborative filtering, and hybrid methods.
- Analyze the accuracy of models using the Root Mean Square Error (RMSE).
- Determine which models provide the most helpful recommendations.
    - Create a receiver operating characteristic (ROC) Curve.
    - Use Average Precision and Average Rank for comparisons.
- See ‘Possible Future Project’ section below for additional information.

Scope of solution space

This project will only focus on learning the concepts and processes of creating a recommendation engine. Deploying it or combining it into a full stack web application will not be part of the second capstone. See ‘Possible Future Project’ section below for additional information.

Constraints

Time is the main constraint because there are many different models and methods that can be used. Will discuss various possibilities during weekly mentor meetings to get feedback as needed to help address this issue.

Stakeholders

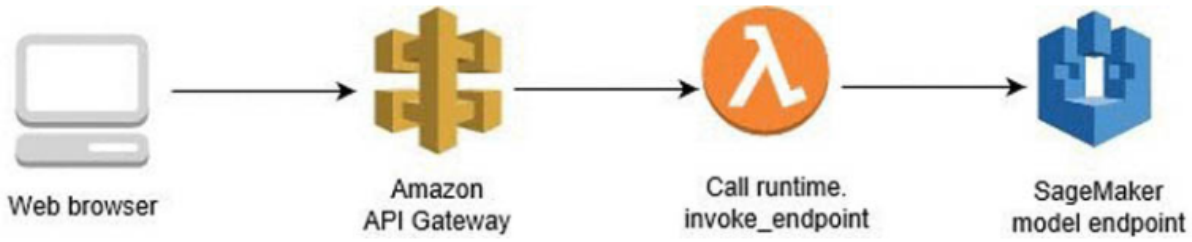
Me for learning purposes.

Data sources

Amazon has a large amount of publicly available customer shopping (implicit feedback) and review data (explicit feedback) from a variety of sources. Will use the [Amazon Review Data (2018)] for this.

Possible Future Project(s)

This project could be deployed into production once the proof of concept was proven successful. One possibility would be to use the following serverless architecture:



Then, on the company's website, do A/B testing to determine the most used recommendations by analyzing Click Through Rate (CTR), Conversion Rate (CR), and Return on Investment (ROI).

Finally, the models(s) and the related recommendations would continue to be enhanced by adjusting to new data, the user's behavior, and purchases over time.

Resources

Jianmo Ni, Jiacheng Li, Julian McAuley. Empirical Methods in Natural Language Processing (EMNLP) (2019). Justifying recommendations using distantly-labeled reviews and fined-grained aspects. Available from: Amazon review data.

Muffaddal Qutbuddin (2020). Available from: An Exhaustive List of Methods to Evaluate Recommender Systems. How to evaluate a recommender system using different evaluation metrics.

Rumi Olsen (2018). Available at: Call an Amazon SageMaker model endpoint using Amazon API Gateway and AWS Lambda | Amazon Web Services .

Tomáš Řehořek (2016). Available from: Evaluating Recommender Systems: Choosing the best one for your business.