DS 223: Marketing Analytics

Group Project: Milestone 1

Anna Charchyan
Alexander Shahramanyan
Yeva Manukyan
Lilith Asminian
Maria Petrosyan

Problem Definition

Identify the Problem Area:

The problem area falls within the domain of Customer Satisfaction and Retention in the retail industry, particularly in the context of product unavailability. When a customer's desired product is out of stock or unavailable, it can lead to lost sales and disappointed customers. Addressing this issue can help retailers enhance customer satisfaction, increase sales, and foster customer loyalty.

Conduct Preliminary Research:

Research shows that recommender systems can significantly improve customer satisfaction and loyalty in the retail sector. Accurate and diverse recommendations based on customer preferences and product attributes are crucial.

Surveys indicate that a majority of readers are open to trying new authors and genres based on recommendations from trusted sources. This presents an opportunity for bookstores to introduce customers to new books they might enjoy when their initial preference is unavailable.

These findings suggest that implementing a book recommendation system can help address the challenges of customer satisfaction and retention by providing personalized suggestions and helping customers discover new books they may like.

Define a Specific Problem

When a customer requests a book that is currently unavailable in a bookstore, there is a risk of losing the sale and potentially disappointing the customer. This problem can be addressed by developing a book recommendation system that suggests similar books based on attributes such as genre, author, and book description. By

providing personalized recommendations, bookstores can increase the likelihood of a customer making a purchase, even if their initial preference is not in stock, thereby improving customer satisfaction and retention.

Methodology

Data Collection

- We will search for and acquire a suitable online dataset of books that includes information such as titles, authors, genres, and detailed book descriptions.
- The dataset should be comprehensive, covering a wide range of books across various genres and authors to ensure diverse recommendations.

Analytical Techniques

- **Text Embedding:** We will use text embedding techniques to convert the book descriptions into dense vector representations. These embeddings will capture the semantic meaning and context of the words in the descriptions.
- **Genre Encoding:** We will also incorporate the genres into the recommendation model by either filtering by genre based on user preferences or calculating genre similarity between books. This will allow the model to consider both content similarity and genre preferences when generating recommendations.
- **Similarity Measurement:** We will calculate the similarity between books based on their description embeddings using cosine similarity or other distance metrics. The similarity scores will be used to identify books with similar content.
- **Recommendation Generation:** When a customer's requested book is unavailable, we will retrieve the most similar books based on the calculated similarity scores and genre preferences, and recommend them to the customer.

Implementation Plan

- **Data Preprocessing:** We will preprocess the acquired dataset by cleaning, transforming, and integrating it into a suitable format for analysis. This includes embedding the book descriptions and encoding the genre information.
- **Recommendation Engine:** We will develop a recommendation engine that generates book recommendations based on content similarity and genre preferences. When a customer's requested book is unavailable, the engine will perform the following steps:

    - Retrieve the description and genre information of the requested book.
    - Calculate the similarity scores between the requested book and other books in the dataset.
    - Filter or rank the books based on their similarity scores and genre preferences.
    - Return the top-N most similar books as recommendations.
- **API Creation:** We will create a well-documented API that exposes the functionality of the recommendation engine. The API will accept requests containing information about the unavailable book and return the recommended books as a response. This will allow bookstores to easily integrate the recommendation system into their existing platforms.
- **Bookstore Collaboration:** We will reach out to bookstores and establish partnerships to test our recommendation system in real-world scenarios. This collaboration will allow us to gather valuable feedback and data on the system's performance and user satisfaction.
- **Evaluation and Refinement:** We will evaluate the effectiveness of the recommendation system using the Successful Recommendation Rate, which measures the percentage of recommendations that resulted in a sale. Based on the evaluation results, we will iteratively refine the recommendation engine. Additionally, we also consider feeding the data on recommendation results back into the model, allowing it to improve its performance.

Expected Outcomes

Implementing the book recommendation system is expected to bring several benefits to bookstores. By providing recommendations when a requested book is unavailable, the system aims to increase customer satisfaction, boost sales and revenue, enhance customer loyalty, and improve inventory management. The system will help convert potential lost sales into successful purchases.

Evaluation Metrics

The success of the book recommendation system will be primarily measured using the Successful Recommendation Rate (SRR) as discussed earlier. The SRR calculates the percentage of recommendations that result in a customer making a purchase, directly measuring the system's effectiveness in generating relevant and appealing recommendations. By monitoring and analyzing the SRR, bookstores can assess the system's impact on their business and make data-driven decisions to optimize its performance over time.